



UNIVERSITAT_{DE}
BARCELONA

Development and neural bases of the spatial recoding of acoustic pitch

Irune Fernández Prieto



Aquesta tesi doctoral està subjecta a la llicència **Reconeixement- Compartlqual 3.0. Espanya de Creative Commons.**

Esta tesis doctoral está sujeta a la licencia **Reconocimiento - Compartirlqual 3.0. España de Creative Commons.**

This doctoral thesis is licensed under the **Creative Commons Attribution-ShareAlike 3.0. Spain License.**



UNIVERSITAT DE
BARCELONA



Development and neural bases of the spatial recoding of acoustic pitch

This thesis is presented by:
Irene Fernández Prieto

Supervisors:
Dr. Jordi Navarra and Dr. Ferran Pons

Doctoral program in Personality and Behaviour
Sant Joan de Déu Foundation
Department of Cognition, Development and Educational Psychology, University of
Barcelona

A mi hermana Irantzu

A Amatxu y a Aita

*“Sometimes science is more art than science,
a lot of people don’t get that”*

Dr. Rick Sánchez

ACKNOWLEDGMENTS

Ya estamos en la recta final, y no me queda más que decir que gracias.

Primero de todo mi inmenso agradecimiento y cariño a Jordi y Ferran por la paciencia de dirigirme y enseñarme todos estos años. Yo he tenido la suerte de tener dos magníficos “jefes”, buenos científicos y buenas personas. Los dos han tenido el “don” de complementarse, aun siendo estilos muy diferentes, y les estaré eternamente agradecida porque eso ha hecho que yo haya podido disfrutar del camino. Si volviera hacia atrás, hubiera elegido a los mismos supervisores, eso lo dice todo.

A mis “labs” donde no he tenido compañeros, sino amigos. A Laura ¿qué puedo decir? Gracias por las risas (muchas, demasiadas para mis endeble abdominales), por enseñarme a apreciar el papel satinado (igual apreciar es una palabra muy fuerte), por tu compañía (la única en más de una ocasión). Aunque bebas leche de soja siempre tendrás mi amistad, nadie es perfecto. Esta tesis es muy tuya también (literalmente). Claudia, te voy a echar de menos. Gracias por todo, por ser un ejemplo de tenacidad y por intentar hacer de mí una montañera en los primeros años, casi lo consigues. Mejor seguimos quedando en algún bar para comer carne en salsa, no forcemos la máquina. Sara, fue corto pero intenso (desafortunadamente para mí), pero seguimos colaborando y pienso volver a Liverpool con alguna excusa “académica” para veros a ti y a Carlos. A los compañeros y amigos del APAL (Jorgina, Maria, Helena, Itxi, Anna, Joan, Clem, Jessica, Marta... espero no dejarme nadie), que buenos ratos hemos compartido. Y también al lab de oncología por adoptarme más de una vez: Nagore, Guillem y Laura.

I'd like to thank Charles (Spence) to bring me the opportunity to join his lab in Oxford, it was one of the best experiences of my academic life, and my personal life. Gracias Alejandro, por cuidarme y enseñarme en Oxford, y por brindarme un futuro excitante, te echo mucho de menos, y a María que eras la alegría de la huerta (no sé si eso se dice en Colombia). Also, I'd like to thank my labmates and friends Noemi and Farrah.

Thanks to Athena (Voloumanos) for inviting me to her lab in New York, it was a wonderful experience. Thank you to my labmates in NY too.

Ruggero e Alice, senza di voi tutto questo non sarebbe stato lo stesso. Voi mi conoscete, non mi piace cambiare, vorrei tanto mettervi in una valigia e portarvi dall'altra parte del mondo, così potrete fare il caffè e gli Espaguetis (con l'accento sulla E). Grazie mille. Già!-lo? E dal momento che sto scrivendo in italiano (grazie Gava per correggere il mio orribile italiano), grazie anche ai miei amici di Oxford, per avermi accettato nel loro ghetto: Luca, Marco (ce l'hai fatta!), Dario, Ale, Michele, Giulio e Ludo (merci!), un grande abbraccio a tutti. I a Albert, t'estimo bro!

Amaia, Mabel y Lidya, ¿Qué hacemos ahora? No hay suficiente espacio para deciros lo importantes que habéis sido y sois en mi vida. Eskerrik asko. Y gracias a Fran (Pako) por seguir teniendo sed, a Oskar por su alma de rockero, a Alex (y a las crías) porque desde el primer día fue/fueron familia, y a Miriam, por esos domingos de budweiser, muchas gracias.

A la gente de la UPF, en especial a Ramon, Phillip, Alex, Iñigo y Rubén. Gracias, muchas gracias por tantas y tantas comidas en el "hospi". Se os echa de menos.

También no me olvido de mis compañeros y amigos de la UIB. Un agradecimiento especial a mis profesores y supervisores de carrera y máster, a Marcos (Nadal), Enric (Munar) y Toni (Gomila) por introducirme en el mundo de la investigación. Y a Josu, David y Rosa, mi familia de Mallorca. Y a mi gente de Ibiza: Ali, Marta y Bea, después de 27 años de amistad. Gracias por seguir ahí.

También me gustaría extender mi serio y rotundo agradecimiento a: los creadores de Microsoft Paint, Terry Pratchett, la moda de los “gin tonics”, el Deezer, el menú subvencionado del hospital, el bar Makinavaja, la grapadora del APAL, las tortitas de maíz con sabor a cartón (que han salvado de la hambruna a este laboratorio en varias ocasiones), Archer, los mejillones en escabeche, la canción “La gozadera”, el bar Spritz, Game of thrones, Oxford Backpackers, la línea 3 del metro de Barcelona, el pub Port Mahón de Cowley, el “loser móvil” y el café instantáneo de Nescafé (que está asqueroso pero ha facilitado la redacción de esta tesis).

Y por último, a mi familia. A mis aitas que me han apoyado ciegamente y confiado en mí. Y para finalizar a mi hermana Irantzu, que hace ya casi 10 años cuando dejé la carrera de física me dijo: - ¡Enana, creo que deberías estudiar Psicología!, a lo que yo le contesté: - No. Os podéis imaginar quien ganó la discusión.

Al final nos hemos divertido, eso es lo importante.

A todos y cada uno de vosotros.

Gracias, Grazie, Gràcies, Eskerrik asko, Danke, Merci and Thanks.

irune

“Al lugar donde has sido feliz no debieras tratar de volver”

Joaquín Sabina. Canción: Peces de Ciudad

SUMMARY

Previous studies suggest the existence of crossmodal correspondences between pitch elevation and visuospatial dimensions (e.g. spatial elevation or visual size) (e.g., Gallace & Spence, 2006; Rusconi et al., 2006). Neuroimaging research has revealed the activation of parietal areas of the brain (e.g., the intraparietal sulcus, IPS), traditionally associated to spatial functions, during the performance of pitch-based musical tasks (see Foster & Zatorre, 2010a, 2010b). This dissertation aimed to investigate the cognitive, developmental and the possible neural mechanisms underlying these crossmodal correspondences.

In the first part of the dissertation, we explored the role of crossmodal correspondences in the modulation of visuospatial attention. We explored how a specific dynamic sound (e.g., an ascending sound) facilitates exogenously the spatial processing of a visual stimulus in adults. For this purpose, we used a modification of the Posner paradigm where ascending and descending frequency sweeps were used as spatial cues instead of the traditional visually-presented cues.

In a second part of the dissertation, we explored the development of the possible mechanisms involved in the crossmodal correspondence processing. More specifically, we investigated whether these mechanisms are present from very early on, or they appear as the experience with the environment increases. To do so, we explored the association between pitch-elevation and visual size in 4- and 6-month-old infants.

Finally, in an attempt to tentatively establish the functional role of certain areas of the right parietal lobe in the association between pitch and spatial elevation, we analysed whether a neurological syndrome associated to structural and functional anomalies of the right parietal lobe (the Nonverbal Learning Disorder, NLD) could affect the performance of pitch-based tasks. To achieve this goal, we compared the performance in pitch-based and spatial tasks in patients with NLD and in a control group with no psychiatric or neurologic disorders.

The results of this three studies that constitute this dissertation suggested: (1) that ascending frequency sweeps exert a stronger modulation of the perceptual system's reaction to upcoming visual targets than descending frequency sweeps; (2) the existence of crossmodal associations between pitch and size in 6-month-old but not in younger infants, suggesting that experience and/or further maturation is needed to fully develop this association; and (3) that functional or structural disruption in parietal areas has a negative impact on auditory processing tasks such as discriminating pitch-elevation differences.

In conclusion, I demonstrated that auditory pitch modulates spatial attention. Secondly, crossmodal correspondences are controlled by basic mechanisms that are present from very early on. Finally, I showed evidence that the perception of complex auditory information may recruit areas of the brain that are compromised in neurologic syndromes that affect the processing space.

RESUMEN

Estudios previos sugieren la existencia de correspondencias transmodales entre la altura tonal y las dimensiones visuoespaciales (por ejemplo, la elevación espacial o el tamaño visual) (Gallace y Spence, 2006; Rusconi y cols., 2006). La investigación en neuroimagen ha puesto de manifiesto la activación de áreas parietales (por ejemplo, el surco intraparietal, IPS) tradicionalmente asociadas a funciones espaciales, durante la realización de tareas musicales relacionadas con la altura tonal (Foster y Zatorre, 2010a, 2010b). El objetivo de esta tesis doctoral fue investigar los mecanismos cognitivos, de desarrollo y neurales que subyacen a estas correspondencias transmodales.

En la primera parte de la tesis, se estudió el papel de las correspondencias transmodales en la modulación de la atención. Exploramos cómo un sonido dinámico específico (por ejemplo, un sonido ascendente) modula exógenamente la localización espacial de un estímulo visual en un adulto. Los participantes realizaron una tarea que consistió en una modificación del paradigma clásico de Posner, en la que usamos sonidos ascendentes y descendentes como señales auditivas.

En una segunda parte de la tesis se analizó el desarrollo de los mecanismos implicados en el procesamiento de las correspondencias transmodales. Más específicamente, se investigó si estas están presentes desde muy temprano en el desarrollo, o aparecen debido a al incremento de la experiencia en el entorno. Para ello, se observó la asociación entre la altura tonal y el tamaño visual en bebés de 4 y 6 meses de edad en un paradigma audiovisual de preferencia visual.

Por último, en un intento por establecer el papel funcional de ciertas áreas del lóbulo parietal derecho en la asociación entre la altura tonal y la elevación espacial, analizamos si los síndromes neurológicos, asociados con una alteración funcional o estructural del lóbulo parietal derecho (por ejemplo, el trastorno de aprendizaje no verbal, TANV), pueden afectar al desempeño de tareas relacionadas con la altura tonal. Para lograr nuestros objetivos, se comparó el rendimiento en tareas

visuoespaciales y auditivas de pacientes con pacientes con TANV y un grupo de control sin trastornos neurológicos o psiquiátricos.

Los resultados de estos tres estudios que constituyen esta tesis sugieren: (1) que los sonidos ascendentes pueden ejercer mayor modulación del tiempo de reacción del sistema perceptivo respecto a los objetos visuales que los sonidos descendentes, (2) la existencia de asociaciones transmodales entre la altura tonal y el tamaño visual en bebés de 6 meses de edad, pero no en bebés más pequeños, lo que sugiere que la experiencia y/o maduración es necesaria para desarrollar plenamente esta asociación, y (3) que la alteración funcional o estructural de áreas parietales tiene un impacto negativo en las tareas de procesamiento auditivo, por ejemplo, discriminar alturas tonales.

En conclusión, se confirmó que la altura tonal puede modular la atención espacial. En segundo lugar, las correspondencias transmodales están controladas por mecanismos básicos que están presentes desde temprana edad. Por último, también se encontró evidencia de que la percepción de la información auditiva compleja podría requerir áreas del cerebro relacionadas con trastornos neurológicos que afectan el procesamiento del espacio.

DECLARACIÓN

Esta disertación doctoral contiene trabajo original.

Artículos publicados:

Fernández-Prieto, I. & Navarra, J. (accepted, pending minor changes) Spatial encoding of pitch in frequency sweeps. *Psychology of Music*.

Fernández-Prieto, I., Navarra, J., & Pons, F. (2015). How big is this sound? Crossmodal association between pitch and size in infants. *Infant Behavior and Development*, 38, 77-81.

Fernández-Prieto, I., Caprile, C., Ristol-Orrriols, B., López-Sala, A., Póo-Argüelles, Tinocco-Gonzalez, D., P., Pons, F. & Navarra, J. (2016). Auditory deficits in nonverbal learning disability. *Research in Developmental Disabilities*, 59, 378-386.

ÍNDICE DE CONTENIDOS

PREFACE	1
1. INTRODUCCIÓN GENERAL	3
1.1. Introducción a las correspondencias transmodales	3
1.1.1. Asociaciones naturales a través de las modalidades sensoriales	8
1.1.2. Correspondencias transmodales <i>vs.</i> Sinestesia	10
1.2. Recodificación espacial de la altura tonal y las dimensiones visuoespaciales	11
1.2.1. Asociaciones transmodales entre la altura tonal y la ubicación espacial	11
1.2.2. Asociaciones transmodales entre el tamaño y la altura tonal	16
1.2.3. Influencia del procesamiento transmodal en la atención	17
1.2.4. Asociaciones transmodales en la infancia y la niñez	21
1.2.5. Bases neurales de la recodificación espacial de la altura tonal	23
2. OBJETIVOS	29
3. STUDY 1: Spatial encoding of pitch in frequency sweeps	31
4. STUDY 2: How big is this sound? Crossmodal association between pitch and size in infants	52
5. STUDY 3: Auditory deficits in preadolescents and adolescents with nonverbal learning disorder	64
6. DISCUSIÓN	89
6.1. Modulación de la atención a través de asociaciones transmodales en adultos	90
6.2. Asociaciones transmodales en bebés prelingüísticos	93
6.3. Procesamiento auditivo en pacientes con alteraciones visuoespaciales	97
7. CONCLUSIONS	100
8. FUTURAS DIRECCIONES	101
9. REFERENCIAS	103

*“All natural arrangements,
however capricious they may seem, have a function.”*

Santiago Ramon y Cajal

PREFACE

We live in a world where our brain is constantly stimulated by a huge amount of stimuli. Our brain receives and interprets efficiently all this information in a way that any artificial intelligence (at least so far) could process.

Our sensory systems cannot be understood as isolated units. Most of the tasks performed by our brain are based on the integration of signals from different modalities. Thus, the existence of crossmodal associations (so-called crossmodal correspondences) between sensory features (e.g., size or pitch) from different modalities should not surprise us, as they may help us to both to simplify and facilitate information processing.

The main goal of the present dissertation is to investigate how crossmodal correspondences can influence certain cognitive processes in adults, and how factors such as experience and cognitive and/or neuronal development can modulate these crossmodal effects. Furthermore, in order to better understand the neural bases of these associations between sensory modalities, this research also focused on how the disruption in a sensory modality can affect the correct functioning in another different sensory modality.

This doctoral dissertation is divided into nine sections. In section 1, my aim is to provide a theoretical background of the current state of the research related to crossmodal correspondences and the possible spatial recoding of acoustic pitch. In Section 2, I describe the main objectives of the research included in this doctoral dissertation. In Section 3, I present the first experimental study of this doctoral dissertation. In this experimental work, I aimed at investigating if high-pitched sounds are associated to a higher location in the space, modulating the visual spatial attention. I also analyzed the possible influence of arousal in this effect. In Section 4, I introduce the second experimental study, where I investigated the crossmodal

association between pitch-elevation and visual size from a very early stage of development. In Section 5, I report the third experimental study focused on the impairments in auditory frequency discrimination due to a cognitive disorder (the so called nonverbal learning disability) that mainly affects visuospatial skills. In Section 6, I discuss the results obtained in the three studies included in the present doctoral dissertation. In Sections 7 and 8, I provide final the conclusions that can be drawn based on my experimental work and suggest some future directions, respectively. And finally, in Section 9, I present the general references of this doctoral dissertation.

1. Introducción General

1.1 Introducción a las correspondencias transmodales

La integración multisensorial se describe como un proceso que permite a nuestro cerebro y al de otros animales integrar la información que llega a las diferentes modalidades sensoriales (por ejemplo, tacto, vista, gusto, etc.).

La interacción entre diferentes modalidades sensoriales posibilita a nuestro sistema cognitivo crear un percepto unificado y coherente de toda la información percibida. En realidad, es difícil imaginar un evento donde podamos experimentar un estímulo aislado en una sola modalidad sensorial sin que otro o más sistemas sensoriales procesen un determinado grado de información del estímulo (Bayne y Spence, 2015).

Normalmente la integración proviene de diferentes señales sensoriales que pertenecen a un mismo objeto externo, por ejemplo, el movimiento de la boca (estímulo visual) y el sonido del habla (estímulo auditivo). Diversos estudios han demostrado que la percepción de señales redundantes de un mismo objeto o evento mejoran su procesamiento (Ernst y Bühlhoff, 2004; Trommershäuser, Landy, y Kording, 2011).

Sin embargo, en otras ocasiones, ambas señales sensoriales pueden pertenecer a diferentes objetos externos y, aun así, nuestro sistema perceptivo puede procesar esta información multisensorial como un único evento. Por ejemplo, en la ilusión de ventriloquia, un ventrílocuo puede hablar sin mover visiblemente los labios mientras manipula una marioneta en sincronía con su discurso. Así, el ventrílocuo crea la ilusión de que las palabras provienen de la boca de la marioneta, es decir, se produce un sesgo perceptivo en el cual el sonido es percibido desde una dirección distinta a su fuente verdadera debido a la influencia e integración con la información visual (Bertelson y de Gelder, 2004).

Durante la mayor parte del siglo XX, gran parte de la investigación experimental se caracterizó por tener un enfoque unimodal que monopolizó el estudio de la percepción. La percepción sensorial ha sido tradicionalmente estudiada como

diferentes modalidades sensoriales aisladas que operan con mecanismos encapsulados y separados entre sí. Sin embargo, a partir de la década de los 70 hubo un aumento de estudios experimentales con un enfoque multimodal (Calvert, Spence, y Stein, 2004). Según esta perspectiva multimodal, la percepción es el resultado de un proceso de interacción e integración entre distintas modalidades sensoriales.

La integración multisensorial de la información ha sido extensamente estudiada desde una perspectiva neural y/o neuronal (Stein y Stanford, 2008). La combinación de información de dos modalidades distintas (por ejemplo, visión y audición) puede evocar una respuesta mayor que la presentación aislada de dicha información en cada modalidad por separado, es decir, una respuesta superaditiva¹.

Meredith y Stein (1983) observaron mayor actividad neuronal en el colículo superior (estructura cerebral situada en el mesencéfalo, ver Figura 1) ante la exposición simultánea a estímulos visuales, auditivos y/o somatosensoriales en gatos (para estudios con humanos, véase Calvert, Hansen, Iversen, y Brammer, 2001). Estos estudios postularon que el máximo nivel de integración multisensorial es más probable que ocurra de acuerdo con tres reglas o leyes (Stein y Meredith, 1993): la regla temporal, la regla espacial, y la ley de la eficacia inversa. En la regla temporal los estímulos unimodales tienen que ser presentados aproximadamente al mismo tiempo (King y Palmer, 1985; Meredith, Nemitz, y Stein, 1987). La regla espacial hace referencia a estímulos unimodales provenientes de aproximadamente la misma ubicación (King y Palmer, 1985; Meredith y Stein, 1986, 1996). La ley de la eficacia inversa se define por una mayor eficacia de integración multisensorial cuando las respuestas unimodales son relativamente débiles (Meredith y Stein, 1983). Además, como se verá a continuación, la integración multisensorial puede verse influenciada por otros fenómenos como las correspondencias transmodales.

¹ La respuesta multisensorial superaditiva es superior a la respuesta aditiva, es decir, a la suma de la combinación de las respuestas a los estímulos unisensoriales

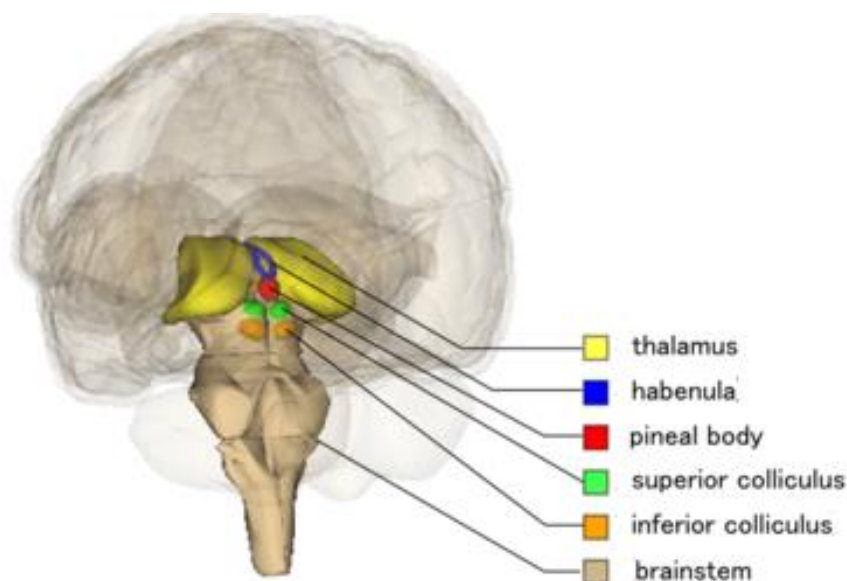


Figura 1. Imagen del tronco cerebral y el tálamo desde la parte posterior. En verde el colículo superior donde las neuronas multisensoriales se encuentran significativamente en mayor proporción respecto a otras áreas del cerebro. (Adaptado con permiso de la imagen original: Life Science Databases (LSDB), CC BY-SA 2.1, https://commons.wikimedia.org/wiki/File:Brainstem_and_thalamus_ja_ja_135.png)

Correspondencia transmodal (también conocida como asociación transmodal) es el término utilizado para describir la tendencia a asociar dos atributos perceptivos distintos, como la altura tonal con la elevación espacial (Spence, 2011). Por ejemplo, al escuchar una melodía musical no solo estamos percibiendo un conjunto de estímulos procesados a nivel auditivo (una única modalidad sensorial), sino que probablemente, también a nivel visuoespacial. Es decir, nuestro sistema cognitivo es capaz de situar sonidos en el espacio (arriba o abajo) o dotarlos de cualidades espaciales (redondeados o puntiagudos) (Salgado-Montejo y cols., 2016).

Probablemente la asociación entre audición y visión ha sido la más investigada (Deroy y Spence, 2013; Deroy, Fernández-Prieto, Navarra y Spence, en prensa; Spence, 2011). Aun así, también existen estudios sobre otras asociaciones como por ejemplo la vinculación entre la audición y otras modalidades sensoriales. Por ejemplo, Belkin, Martin, Kemp, y Gilbert (1997) observaron asociaciones entre la altura tonal y el olfato. Los participantes tendían a catalogar tonos específicos con olores específicos. La asociación entre la altura tonal y el sabor también fue mostrada

en un estudio posterior de Crisinel y Spence (2009). En este caso los participantes asociaron el sabor ácido y el sabor amargo con tonos agudos y tonos graves, respectivamente. En cuanto a posibles asociaciones entre la visión y el tacto, Martino y Marks (2000) mostraron que una alta luminosidad se asocia a una alta frecuencia vibrotáctil y una baja luminosidad se asocia con una baja frecuencia vibrotáctil.

Observadas en conjunto, parece que las asociaciones transmodales podrían surgir entre todas las combinaciones de modalidades sensoriales. Sin embargo, debido a que, la presente disertación estudia las asociaciones transmodales audiovisuales, a partir de ahora nos centraremos únicamente en el vínculo auditivo y visuoespacial.

Uno de los ejemplos más conocidos de asociación transmodal audiovisual es el llamado efecto Bouba / Kiki (Ramachandran y Hubbard, 2001), en el cual el sonido de una palabra es asociado a una figura en concreto. En una réplica del estudio original de Köhler (1929), Ramachandran y Hubbard (2001) encontraron una asociación transmodal audiovisual entre el sonido de unas palabras y unas figuras geométricas. En el estudio se pidió a los participantes nombrar dos figuras con la etiqueta léxica “bouba” o la etiqueta léxica “kiki”. El 98% de los participantes asoció el sonido de la palabra "kiki" con la figura puntiaguda y el sonido del término "bouba" con la figura redondeada (ver Figura 2).

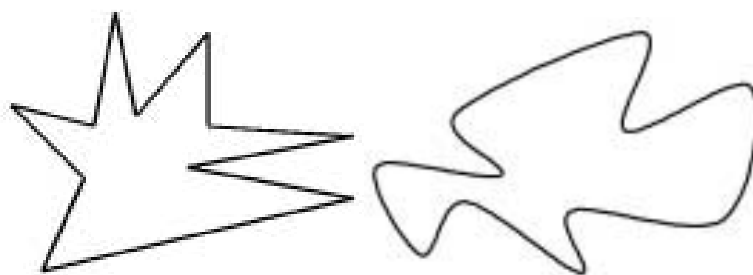


Figura 2. Los participantes asociaron el sonido "kiki" con la figura puntiaguda de la izquierda y el sonido "bouba" con la figura redondeada de la derecha (Ramachandran y Hubbard, 2001, 2003).

El efecto Bouba/Kiki también ha sido investigado en bebés prelingüísticos². Sin embargo, la asociación entre el sonido de los términos léxicos y las figuras visuales no parece ser tan robusta en los primeros años de vida. Ozturk., Krehm y Vouloumanos (2012) reportaron efectos de asociación simbólica en bebés de 4 meses de edad. Los bebés miraron más tiempo a las figuras que se presentaban acompañadas del sonido de una palabra congruente (por ejemplo, figura puntiaguda presentada con el sonido de la palabra /kiki/). Sin embargo, Fort, Weiß, Martin, y Peperkamp (2013) no encontraron efectos de esta asociación simbólica en bebés de 5 y 6 meses de edad. Estos investigadores utilizaron un diseño más complejo y emplearon al menos 6 pares de palabras (por ejemplo, /lumu/ y /lomo/ o /kiti/ y /tiki/). Fort y colaboradores (2013) argumentaron que dichos resultados contradictorios se deberían a la necesidad de diseños más sencillos para detectar el efecto Bouba-Kiki, probablemente debido a una posible falta de solidez de este fenómeno en bebés prelingüísticos. En contraste, Maurer, Pathman y Mondloch (2006) mostraron que los niños de 2,5 años ya muestran un efecto Bouba/Kiki similar a los adultos. Estos últimos resultados sugieren la importancia del desarrollo del lenguaje para la emergencia de este fenómeno.

El efecto Bouba/Kiki nos lleva a plantearnos si estas asignaciones audiovisuales podrían no ser tan arbitrarias como podríamos suponer en un principio, y estos patrones se habrían desarrollado debido a la influencia de las correspondencias transmodales entre el sonido de las etiquetas léxicas y la forma visual. Ramachandran y Hubbard, (2001), argumentaron que este fenómeno podría emerger debido a la asociación entre la percepción visual de la figura (redondeada o puntiaguda) y la forma que presentan los labios al producir los fonemas de las palabras (por ejemplo, una apertura de la boca más redondeada para el término /bouba/ y una apertura de la boca menos redondeada para el término /kiki/). Debido a ello, no se puede asumir que la asignación de etiquetas léxicas a objetos es arbitraria en las diferentes lenguas, como evidencia, también existen las

² La fase de prelingüística en los bebés es el período comprendido entre los 0 y los 13 meses, esta etapa es anterior a la pronunciación de sus primeras palabras. En la etapa prelingüística, la mayoría de los sonidos que produce el bebé no son más que práctica de la manipulación y la secuenciación del sonido (Owens, 2008).

onomatopeyas dónde el sonido está directamente relacionado con el objeto (por ejemplo, la onomatopeya "*boom*" como la analogía de un sonido de explosión).

Las asociaciones entre ciertas dimensiones auditivas y visuoespaciales no parecen crearse de forma arbitraria. La notación musical (representación de sonidos en un pentagrama) es un buen ejemplo de cómo la mayoría de culturas comparten los mismos patrones de correspondencias transmodales.

1.1.1 Asociaciones naturales a través de las modalidades sensoriales

Determinadas asociaciones transmodales parecen emerger de forma sistemática y universal. El origen probablemente universal de la asociación entre la altura tonal y la elevación espacial ha sido evidenciado por hallazgos arqueológicos que datan de la época de los antiguos griegos. El epitafio de Seikilos, que es la composición musical completa más antigua descubierta hasta el momento, mostraba los ascensos en la frecuencia del sonido en posiciones espaciales superiores. La columna de la Seikilos fue hallada en Aydin (Turquía) y data aproximadamente del año 100 d.C. Esta representación visual de la música se puede ver tanto en la cultura occidental como en la oriental, por lo tanto. Es altamente probable, por tanto, que el sistema de notación musical se base en representaciones y asociaciones universales. Estos descubrimientos sugerirían que las correspondencias transmodales entre el la altura tonal y la ubicación espacial tendrían una prevalencia común en la población, en consecuencia, no emergerían debido a factores culturales.

Estos patrones musicales podrían reflejar las asociaciones transmodales entre la localización espacial en el eje vertical y los tonos musicales. Los patrones musicales tienen una cierta equivalencia con la física acústica de nuestro entorno natural: los sonidos que se propagan cerca del suelo tienden a que su espectro de alta frecuencia sea absorbido por diversos objetos (en el medio existe mayor cantidad de objetos a ras del suelo), filtrándose, a ras de suelo, las frecuencias bajas en mayor medida que las frecuencias altas. A causa de este fenómeno, los animales que se mueven a ras de suelo tendrán la tendencia a percibir frecuencias bajas desde localizaciones inferiores,

en el espacio y frecuencias altas desde localizaciones superiores en el espacio. En un estudio reciente, Parise, Knorre y Ernst (2014) encontraron, midiendo directamente los estímulos auditivos distales y proximales³, que las frecuencias altas y bajas tienden a provenir de fuentes localizadas arriba y abajo en el espacio, respectivamente. Como resultado, la exposición repetida a estas correlaciones estadísticas en el entorno natural podría predisponer a nuestros sistemas sensoriales para la adquisición de mecanismos de asociación transmodal.

Evidenciando el posible origen de los patrones de asociación audiovisuales, ya hace más de un siglo el filósofo Carl Stumpf (1883) observó que existía un patrón asociativo en la mayoría de lenguas, las cuales compartían términos como "alto" y "bajo" para definir tanto la altura tonal como la elevación espacial. Recientemente, Parkinson, Kohler, Sievers, y Wheatley (2012) llevaron a cabo un estudio con el fin de investigar la universalidad de la asociación altura-tonal/elevación-espacial. Para ello, seleccionaron un grupo de participantes de una tribu aislada en Camboya, cuyo idioma no utiliza etiquetas lingüísticas espaciales como "alto" o "bajo" para clasificar la altura tonal. Los participantes mostraron tener correspondencias transmodales entre la elevación espacial y la altura tonal como los hablantes de lenguas con términos compartidos para elevación espacial y altura tonal, indicando que el lenguaje no es indispensable para esta asociación. Así pues, aunque la mayoría de las lenguas, como ya afirmó Stumpf (1883), comparten etiquetas lingüísticas para describir la altura tonal y la elevación espacial, éstas no parecen ser necesarias para la creación de determinadas correspondencias transmodales.

En la misma línea, y tal vez con mayor frecuencia y robustez, nos encontramos con la asociación entre el sonido y tamaño. Los objetos más grandes tienden a producir sonidos con frecuencias más bajas (por ejemplo, una tuba o un elefante), mientras que los objetos más pequeños tienden a producir sonidos de frecuencias más altas (por ejemplo, un flautín o un colibrí). La explicación, desde punto de vista de la física,

³ El término estímulo distal hace referencia al objeto o evento que proporciona la información sobre el estímulo proximal. El estímulo proximal viene dado por el procesamiento de la información del estímulo distal a través de los receptores sensoriales. Por ejemplo, cuando escuchamos música el estímulo distal sería las diferentes frecuencias auditivas producidas por el reproductor musical, en cambio, el estímulo proximal sería el procesamiento del sonido por parte nuestro sistema auditivo.

sería que las frecuencias más bajas son las que tienen mayor longitud de onda, y para producir ondas más largas es necesaria una superficie vibratoria (o resonador) de mayor tamaño. En consecuencia, las asociaciones entre la altura tonal y el tamaño se observan con regularidad en el ambiente físico pudiendo producir un patrón estadístico de asociación transmodal a nivel perceptivo entre la altura tonal y el tamaño (Gallace y Spence, 2006; Walker y Smith, 1985).

1.1.2 Correspondencias transmodales y/o sinestesia

Algunos investigadores han postulado que las correspondencias transmodales serían equivalentes a una sinestesia universal (Marino y Marks, 2001).

La sinestesia, o "la unión de sensaciones o sentidos", se caracteriza por la percepción de un estímulo en una modalidad sensorial que evoca una experiencia perceptiva añadida en la misma o diferente modalidad sensorial (Cohen Kadosh, Cohen Kadosh, y Henik, 2007; Melara y O'Brien, 1987). La forma más común de sinestesia es la llamada sinestesia grafema-color (ver Figura 3). Esta forma de sinestesia se caracteriza por la percepción del color asociado con letras o números específicos (Baron-Cohen, Harrison, Goldstein, y Wyke, 1993). Existen muchas otras formas de sinestesia, aunque no han sido tan estudiadas y su incidencia es menor entre población sinestésica, por ejemplo, la chromestesia (asociación de sonidos con colores) (Rizzo y Eslinger, 1989) o la sinestesia auditiva-táctil (sonidos que pueden inducir sensaciones en el cuerpo) (Naumer y van den Bosch, 2009), entre otras.

(a) SINESTESIA

(b) SINESTESIA

Figura 3. (a) Ejemplo de una palabra escrita en negro percibido por una persona no sinestésica (b) Aproximación de cómo una persona sinestésica podría percibir las letras de una palabra escrita en negro.

La principal diferencia entre la sinestesia y las correspondencias transmodales es que la primera es una condición neurológica poco común que varía de persona a persona, mientras que las asociaciones transmodales tienden a seguir patrones universales. La incidencia de la sinestesia es baja, aproximadamente en 1 cada 2000 mujeres y 1 de cada 12000 hombres (Baron-Cohen, Burt, Laittan-Smith, Harrison, y Bolton, 1996). La diferencia en la incidencia entre sexos podría apoyar la hipótesis de que la sinestesia es innata y no aprendida.

Baron-Cohen y colaboradores (1996) demostraron que la sinestesia es un fenómeno que no se deriva de la experiencia con el medio o de la memorización de asociaciones. Pidieron a participantes sinestésicos y a un grupo control describir el color experimentado al escuchar una serie compuesta de palabras y letras. El grupo de control fue informado de que iban a ser reevaluados una semana más tarde. Las respuestas del grupo control fueron consistentes en un 37,6%. Los participantes sinestésicos, a los que no se les informó de la repetición de pruebas, fueron consistentes en un 92,3% con sus respuestas un año más tarde. Estos resultados confirmaron que las asociaciones sinestésicas son constantes en el tiempo.

En resumen, las correspondencias transmodales y la sinestesia difieren en prevalencia, fenomenología y mecanismos cognitivos. En consecuencia, podríamos considerar la sinestesia y la asociación transmodal como dos fenómenos distintos (Martino y Marks, 2001).

1.2. Recodificación espacial de la altura tonal y las características visuoespaciales

1.2.1. Asociaciones transmodales entre la altura tonal y la ubicación espacial

La literatura sobre las correspondencias transmodales contiene numerosos ejemplos de asociaciones entre distintas características sensoriales. La más investigada es la relación entre la altura tonal y las dimensiones visuoespaciales (por ejemplo, la elevación espacial o el tamaño).

Las asociaciones transmodales entre la altura tonal y localización espacial han sido ampliamente documentadas en las últimas décadas. Un elevado número de estudios han demostrado que los sonidos agudos y graves tienden a estar asociados con elevaciones espaciales más altas y más bajas, respectivamente (Mudd, 1963, Pratt, 1930; Occelli, Spence, y Zampini, 2009; Roffler y Butler, 1968; Rusconi, Kwan, Giordano, Umiltà, y Butterworth, 2006; Sonnadara, Gonzalez, Hansen, Elliott, y Lyons, 2009).

Uno de los primeros experimentos realizados para analizar el vínculo asociativo entre la elevación espacial y la altura tonal fue llevado a cabo por Pratt (1930) hace casi un siglo. En su estudio, los participantes tenían que localizar una serie de tonos con distintas frecuencias que eran reproducidos por un altavoz oculto. Los participantes debían localizar los tonos en diferentes lugares de una escala numerada. Los resultados mostraron que los tonos más agudos eran percibidos desde posiciones superiores en el espacio y los tonos más graves desde posiciones inferiores en el espacio. En la misma dirección, décadas después, Mudd (1963) pidió a los participantes representar sonidos mediante el uso de un sistema de clavijas en un panel vertical. Los resultados mostraron que los participantes tendían a colocar las clavijas en las posiciones superiores e inferiores cuando escuchaban una frecuencia más alta o más baja, respectivamente.

Bernstein y Edelstein (1971) fueron quizás los primeros investigadores en estudiar la transmodalidad audiovisual mediante la observación de tiempos de reacción. En la tarea de detección de un estímulo visual, los participantes eran más rápidos cuando el estímulo visual se presentaba en una posición más alta junto a un tono agudo respecto a un tono grave. Los resultados sugieren que la congruencia entre la altura tonal y la elevación espacial podrían producir una discriminación visuoespacial más rápida (Ben-Artzi y Marks, 1995; Evans y Treisman, 2010; Melara y O'Brien, 1987; Patching y Quinlan, 2002).

Siguiendo esta misma línea, Miller (1991) utilizó una tarea Go/No-Go⁴ para demostrar los efectos de congruencia entre la altura tonal y la elevación espacial. En su estudio, un objeto visual aparecía por encima o por debajo de un punto de fijación central y/o se presentaba un tono agudo o grave. Estos estímulos podían ser presentados de manera unimodal, es decir solo la condición visual o la auditiva, o de manera transmodal, es decir, el estímulo visual se presentaba simultáneamente con el tono. En los ensayos Go, en las condiciones visuales los participantes tenían que responder lo más rápido posible a la ubicación del estímulo visual (arriba o abajo) y en las condiciones auditivas debían discriminar el tono (agudo o grave). En los ensayos No-Go, los participantes debían evitar responder cada vez que un estímulo visual se presentaba en el punto de fijación o cuando se presentaba un sonido con un tono intermedio. Los resultados mostraron respuestas más rápidas en los estímulos bimodales congruentes respecto a los incongruentes.

La amplia mayoría de estudios sobre la asociación transmodal entre la altura tonal y la elevación incluyen estímulos auditivos y visuales. En cambio, Ocelli y colaboradores (2009) llevaron a cabo un estudio utilizando dos estímulos auditivos (agudo y grave) y dos estímulos vibrotáctiles (mayor o menor nivel de vibración) que se presentaban conjuntamente. Los participantes debían responder, lo más rápido posible, si el tono presentado correspondía a un tono agudo o a un tono grave. En otros ensayos experimentales, la tarea consistía en decidir si una vibración táctil se presentaba en la parte superior o inferior de un cubo (ver Figura 4). Para responder, en cada bloque (auditivo o táctil) los participantes debían apretar uno de los dos pedales situados a la izquierda o la derecha. Las respuestas de los participantes fueron más rápidas en los bloques congruentes (por ejemplo, vibración en la zona superior junto al sonido agudo) que en los incongruentes (por ejemplo, vibración en la zona superior junto al sonido grave). Estos datos sugieren que existe una asociación entre la altura tonal y la elevación espacial, en este caso percibida a través de la modalidad táctil.

⁴ Las tareas Go/No-Go son utilizadas para medir la atención selectiva y control de respuesta (inhibición) a través del tiempo de reacción y número de errores. En estas tareas los participantes deben realizar una acción en la condición Go (por ejemplo, presionar una tecla) e inhibir la misma acción en la condición No-Go (por ejemplo, no presionar ninguna tecla).

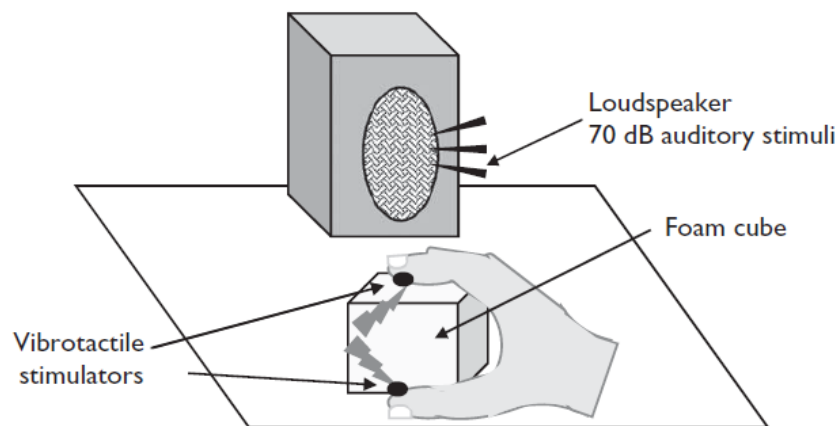


Figura 4. Representación de la configuración y disposición de los estímulos del estudio (Ocelli, Spence, y Zampini, 2009).

La evidencia que sugiere la presencia de la correspondencia transmodal entre la elevación espacial y la altura tonal es bastante sólida, puesto que ha sido observada utilizando distintas metodologías. Rusconi y colaboradores (2006) sugirieron que los efectos de compatibilidad estímulo-respuesta (*Stimulus-response compatibility*, SRC; Proctor y Reeve, 1989) derivados de las correspondencias transmodales serían similares al efecto SNARC (*Spatio-Numerical Association of Responses Codes / Asociación espacio-numérica de códigos de respuestas*), donde el juicio sobre la magnitud de números pequeños es más rápido usando las teclas de la izquierda y el juicio de números mayores usando las teclas de la derecha (Dehaene, Bossini, y Giraux, 1993). Según este planteamiento, los tiempos de reacción más rápidos en las condiciones congruentes respecto las incongruentes podrían ser evidencia de la existencia de una representación mental espacial, es decir, las magnitudes se corresponderían con una distribución espacial de las opciones de respuesta.

En línea con los efectos SNARC y SMARC, Walsh (2003) formuló una teoría de la magnitud (*A theory of magnitude*, ATOM). Walsh hipotetizó que el sistema cognitivo dispondría de un mecanismo común para codificar diversas magnitudes independientemente de la modalidad de entrada. Las magnitudes estarían representadas en la corteza parietal inferior, es decir, el mapeo de cantidades se produciría en coordenadas espaciales como sucede en las correspondencias transmodales entre la altura tonal (magnitud) y la ubicación espacial (coordenadas).

En consecuencia, existiría una población de neuronas en la corteza parietal dedicada a un sistema de representación general de magnitudes temporales, espaciales, numéricas y volumétricas, entre otras.

En un estudio de Rusconi y colaboradores (2006), participantes sin experiencia musical completaron una tarea de discriminación en la que debían comparar la frecuencia de un tono respecto a un tono de referencia pulsando una de las dos teclas situadas en un eje vertical “simbólico” (por ejemplo, "arriba" o "abajo") o en el eje horizontal (por ejemplo, "izquierda" o "derecha") (véase Figura 5). Los resultados mostraron que los tiempos de respuesta se veían afectados la ubicación espacial de respuesta en el eje vertical, pero no en el eje horizontal. Los participantes fueron más rápidos y precisos respondiendo a los tonos agudos y graves cuando la respuesta se correspondía con la localización de la tecla situada "arriba" y "abajo" del teclado, respectivamente. Rusconi y colaboradores definieron este efecto como SMARC (*Spatial–Musical Association of Response Codes / Asociación espacio-musical de código de respuestas*) en analogía con el efecto SNARC. Por otra parte, encontraron efectos de compatibilidad en el eje horizontal solo en músicos expertos, así como efectos de compatibilidad más robustos en el eje vertical que en los participantes no músicos.

Estos datos sugieren que la experiencia musical podría fortalecer el mapeo de la altura tonal en los ejes espaciales. Sin embargo, podría ser que la asociación de la altura tonal con el eje horizontal se debiese a la experiencia musical con el piano (por ejemplo, los tonos agudos se encuentran en las teclas de la derecha del piano). Efectos similares en el eje horizontal fueron observados por Lidji y cols. (2007) en una tarea de discriminación de instrumentos. Los músicos mostraron SRC en el eje horizontal. Sin embargo, la formación en piano en los músicos no correlacionó significativamente con el efecto SRC en el eje horizontal. De acuerdo con estos resultados, el mapeo horizontal de la altura tonal no puede sólo explicarse por la experiencia motora debida a la asociación con la disposición de las notas agudas y graves en cada instrumento musical. Los resultados indican que la experiencia musical interactúa en el efecto SRC, en ambos ejes espaciales: vertical y horizontal.

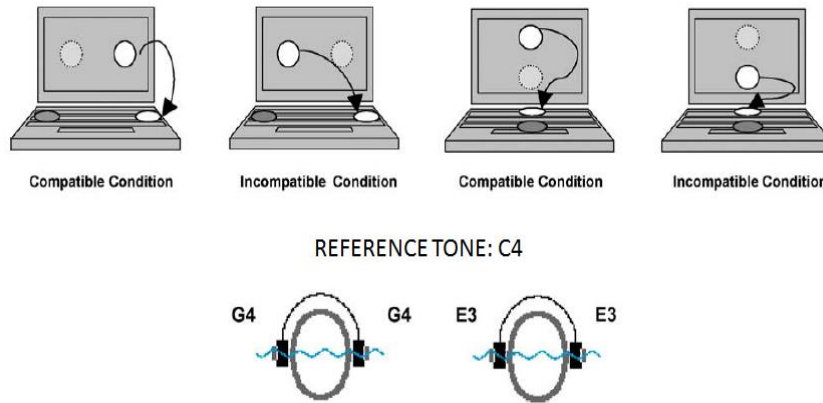


Figura 5. Efectos SRC entre la altura tonal y la posición en el eje horizontal y vertical en las condiciones congruentes e incongruentes. Los estímulos y respuestas correctas se muestran en blanco. (Figura adaptada de Rusconi y cols., 2006).

1.2.2. Asociaciones transmodales entre el tamaño y la altura tonal

A pesar de que no existen tantos estudios, también se han observado efectos similares en la asociación entre la altura tonal y el tamaño. Como hemos mencionado anteriormente, las frecuencias más bajas tienden a ser emitidas en el medio natural por objetos de mayor tamaño, y los sonidos de frecuencia más alta por fuentes de menor tamaño (Parise, Knorre y Ernst, 2014).

Ya en el siglo pasado Sapir (1929) estudió la relación entre el tamaño y el sonido, demostrando que los sonidos del habla (por ejemplo, /a/ - fonema de frecuencia baja - y /i/ - fonema de frecuencia alta-) están asociados con el tamaño de los objetos (por ejemplo, un círculo grande o un círculo pequeño). Los participantes fueron capaces de asociar pseudopalabras como "bal" (sonido con frecuencia acústica más baja) y "bil" (sonido con frecuencia acústica más alta) con objetos más grandes y más pequeños, respectivamente.

Marks, Hammeal, Bornstein, y Smith (1987) usaron una tarea de clasificación de tamaño de un estímulo visual y observaron que los participantes percibían la tarea como más compleja cuando el estímulo visual era presentado junto a un sonido "incongruente" (por ejemplo, un estímulo visual más grande presentado junto con un sonido de frecuencia alta). Siguiendo con esta hipótesis de correspondencia

transmodal entre el tamaño y el sonido, Gallace y Spence (2006) llevaron a cabo una serie de experimentos donde los participantes tenían que juzgar el tamaño de un estímulo visual respecto a un estímulo visual de referencia previo. Cuando un tono se presentaba de manera congruente junto al objeto visual (por ejemplo, mayor estímulo visual, junto con un tono más grave) los participantes mostraban tiempos de reacción más rápidos que cuando el tono presentado era incongruente respecto con el tamaño de la comparación (Bien, ten Oever, Goebel, y Sack, 2012). En conclusión, la frecuencia de un sonido puede influir en la velocidad de procesamiento de un objeto visual, por ejemplo, en la discriminación de su tamaño mediante la asociación transmodal entre dimensiones perceptivas.

1.2.3 Influencia del procesamiento transmodal en la atención visuoespacial

De acuerdo con la creciente evidencia de correspondencias transmodales entre el sonido y las cualidades visuales, ¿en qué medida determinadas frecuencias pueden modular la atención visuoespacial?

El paradigma clásico de atención de Posner (Posner, 1980) fue diseñado originalmente para estudiar la atención visuoespacial. En la tarea de atención exógena, dos recuadros se presentaban en la derecha e izquierda de la pantalla. Uno de los dos recuadros podía parpadear brevemente (señal de localización espacial). Después de cierto tiempo, un asterisco (target visual) podía aparecer en el mismo recuadro que había parpadearado anteriormente (es decir, condición congruente) o en el recuadro opuesto (es decir, condición incongruente) (ver Figura 6). Los participantes eran más rápidos al detectar el asterisco en los ensayos congruentes que en los ensayos incongruentes cuando éste se presentaba después de la señal de localización. Por lo tanto, la detección de un target visual es facilitada por una señal visuoespacial previa que orienta la atención espacial hacia un área específica del campo visual.

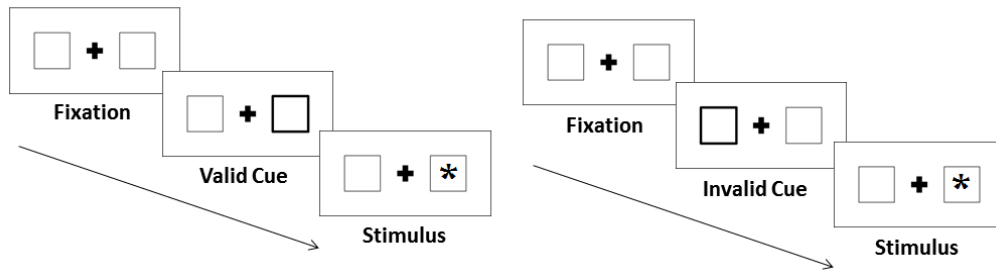


Figura 6. Secuencia de eventos de un ensayo con señales visuales exógenas en el paradigma atencional de Posner para la condición congruente (imagen de la izquierda) y para la condición incongruente (imagen de la derecha). (Imagen adaptada con permiso de Local870, CC BY-SA 3.0, <https://en.wikipedia.org/w/index.php?curid=37904478>)

Diversos estudios sugieren que la atención visuoespacial puede modularse exógenamente mediante el uso de señales previas auditivas, es decir, la señal previa (auditiva) y el target (visuoespacial) de diferentes modalidades sensoriales no presentados al mismo tiempo (Chiou y Rich, 2012; Mossbridge, Grabowecky, y Suzuki, 2011; Spence, 2010). Si la altura tonal puede modular la atención visuoespacial hacia localizaciones superiores e inferiores en el campo visual, deberíamos observar una facilitación en la detección visual de un objeto cuando la señal auditiva y el target visual posterior son congruentes (por ejemplo, una señal auditiva aguda y un objeto visual que aparece posteriormente en una zona de superior del campo visual), demostrando así, que los efectos transmodales ocurren en un nivel previo a la selección de respuesta.

Chiou y Rich (2012) llevaron a cabo una serie de experimentos utilizando modificaciones del paradigma clásico de atención de Posner, con el objetivo de investigar la asociación de los estímulos auditivos en localizaciones espaciales específicas, y en consecuencia, observar como los estímulos auditivos producen un efecto de señalización exógena transmodal que podría modular la detección de estímulos visuales. Para ello, los participantes tenían que detectar un estímulo visual, lo más rápidamente posible, que se presentaba aleatoriamente por encima o por debajo de un punto de fijación central. Antes de la presentación del estímulo visual, se presentaba una señal auditiva irrelevante (por ejemplo, un tono grave o un tono agudo). Los resultados mostraron respuestas más rápidas cuando los participantes

detectaban estímulos visuales en la zona superior e inferior del campo visual precedidos por un tono agudo o grave, respectivamente (ver Figura 5).

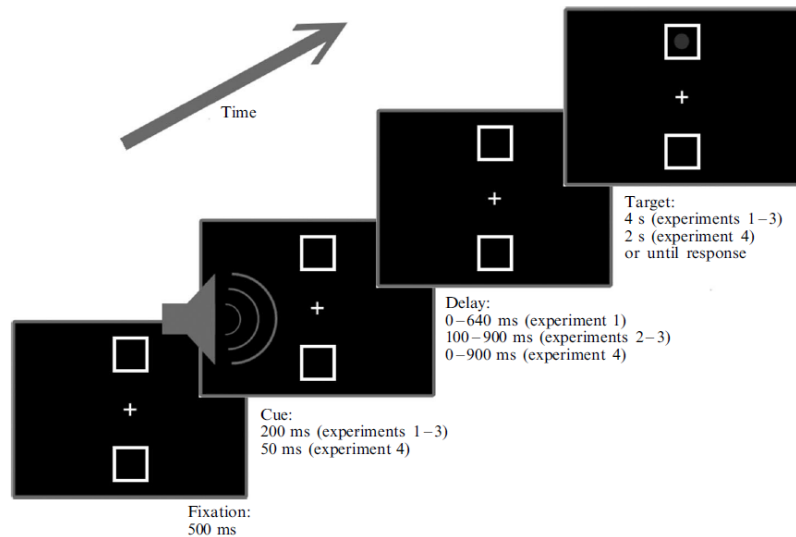


Figura 5. Secuencia de eventos de un ensayo en la condición vertical. En la condición horizontal, las dos cajas estaban situadas a la izquierda ya la derecha de la cruz central de fijación (Chiou y Rich, 2012).

Mossbridge, Grabowecky, y Suzuki (2011) observaron el mismo patrón de respuesta que en el estudio anterior usando una tarea Go/No-Go donde los participantes debían responder cuando el estímulo visual de referencia (un círculo) coincidía con un estímulo posteriormente presentado. El círculo de referencia se presentaba en el punto central de la pantalla conjuntamente con una de dos frecuencias ascendentes o una de dos frecuencias descendentes. El target visual podía presentarse desde cuatro posiciones cardinales (arriba, abajo, izquierda o derecha, Experimento 1, véase Figura 6A) o desde una las cuatro posiciones diagonales respecto al punto de fijación (arriba-izquierda, arriba-derecha, abajo-izquierda, y abajo-derecha, Experimento 2, ver Figura 6B). Los resultados mostraron que los cambios de frecuencia graduales modularon la atención espacial en eje vertical (arriba y abajo). Sin embargo, no se observaron efectos facilitadores en la condición congruente para el eje horizontal (derecha e izquierda) o para las posiciones diagonales.

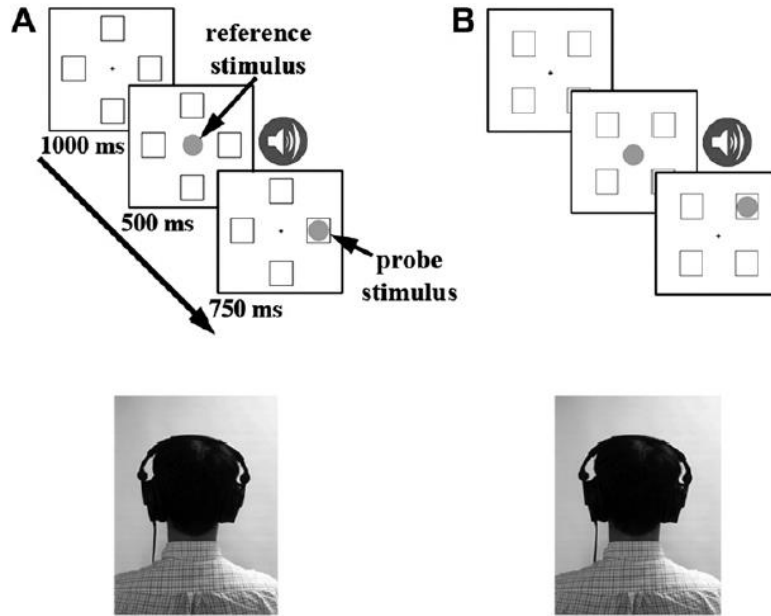


Figura 6. (A) Secuencia de eventos del procedimiento del Experimento 1 de Mossbridge y cols. (2011) con los recuadros situados en los puntos cardinales. (B) Secuencia de eventos del Experimento 2 de Mossbridge y cols. (2011) con los recuadros situados en localizaciones diagonales. (Mossbridge, Grabowecky, y Suzuki, 2011).

Un tema que no ha sido abordado en la literatura hasta este momento, se refiere a si las diferencias perceptivas entre los sonidos graves y agudos afectan a la asociación transmodal entre la altura tonal y la elevación espacial. Las frecuencias altas son perceptivamente más salientes. Una de las posibles explicaciones de este fenómeno es que éstas son procesadas como más intensas que las frecuencias más bajas (Fletcher y Munson, 1933). Kishon-Rabin, Roth, Dijk, Yinon, y Amir (2004) mostraron que es más fácil detectar sonidos ascendentes que sonidos descendentes. Diversas investigaciones, llevadas cabo por Deutsch (1976, 1978) mostraron que, cuando se presentan simultáneamente una frecuencia alta en un oído y una frecuencia más baja en el otro, tendemos a percibir un sonido fusionado en el oído donde se presenta el sonido más agudo. Estos datos sugieren que las frecuencias más altas podrían tener una mayor influencia en la percepción espacial (Deutsch y Roll, 1976; von Békésy, 1963)

A nivel neural, varios estudios utilizando la técnica de electroencefalografía (EEG), han mostrado que los sonidos ascendentes provocan un mayor efecto de *mismatch*

negativity (MNN)⁵ que los sonidos descendentes (Näätänen, 1990; Näätänen, Gaillard, y Mäntysalo, 1978; Ruusuvirta y Astikainen, 2012), evidenciando una mayor influencia de los sonidos agudos, que tienen más saliencia que los sonidos graves, en ciertos procesos atencionales.

De acuerdo con estos estudios, los estímulos auditivos agudos o ascendentes poseen mayores propiedades espaciales inherentes y generan una mayor respuesta electrofisiológica del cerebro. Una pregunta interesante se refiere a si estas propiedades pueden influir en los efectos atencionales que los sonidos agudos (más que los graves) ejercen sobre la detección de estímulos visuales o no.

1.2.4 Asociaciones transmodales en la infancia y la niñez

Como hemos podido ver en los apartados anteriores, existe clara evidencia que apoya la idea de que los humanos adultos asocian múltiples estímulos que llegan a través de diferentes modalidades sensoriales (Spence, 2011). Estas asociaciones transmodales se producen sistemáticamente tal vez debido a la exposición de los patrones asociativos observados en el entorno. Sin embargo, ¿se pueden observar correspondencias transmodales en bebés?

De la misma manera que los adultos, los bebés parecen mostrar correspondencias transmodales, especialmente entre las dimensiones auditiva y el visual (Mondloch y Maurer, 2004; Maurer, Gibson, y Spector, 2012).

Walker y colaboradores (2010), mediante el uso de una tarea de preferencia visual con bebés de 3 a 4 meses de edad, encontraron una tendencia a atender a estímulos auditivos y visuales congruentes. Es decir, los bebés miraban más tiempo a los estímulos visuales que se movían hacia la parte superior de la pantalla (una bola)

⁵ Mismatch negativity (MNN) hace referencia a la respuesta eléctrica mostrada por la electroencefalografía (EEG), que se produce cuando ante la presencia de un estímulo discrepante dentro de una secuencia de estímulos repetitivos (por ejemplo, una secuencia de estímulos auditivos con características acústicas similares)

cuando se presentaban junto con un sonido con una frecuencia ascendente que cuando se presentaba con un sonido con una frecuencia descendente. Dolscheid, Hunnius, Casasanto y Majid (2014) siguieron el mismo procedimiento de Walker y colaboradores (2010) (ver Figura 7) y encontraron resultados similares, es decir, asociaciones transmodales entre frecuencias ascendentes o descendentes y la elevación espacial, así como con la anchura de objetos visuales (por ejemplo, los bebés miraban más tiempo la figura que ensanchaba cuando iba acompañada de un sonido que gradualmente se hacía más grave). Sin embargo, Lewkowicz y Minar (2014) no consiguieron replicar estos efectos mediante la realización de cinco experimentos con bebés de 4, 6 y 8 meses de edad empleando la misma metodología del estudio de Walker y colaboradores (2012). La ausencia de asociaciones transmodales audiovisuales en bebés, reportada en el estudio de Lewkowicz y Minar (2014), podría ser debida a una variación de la intensidad de los estímulos auditivos, que difiere al estudio de Walker y colaboradores (2012) y quizás afectaría a los efectos de transmodalidad audiovisual.

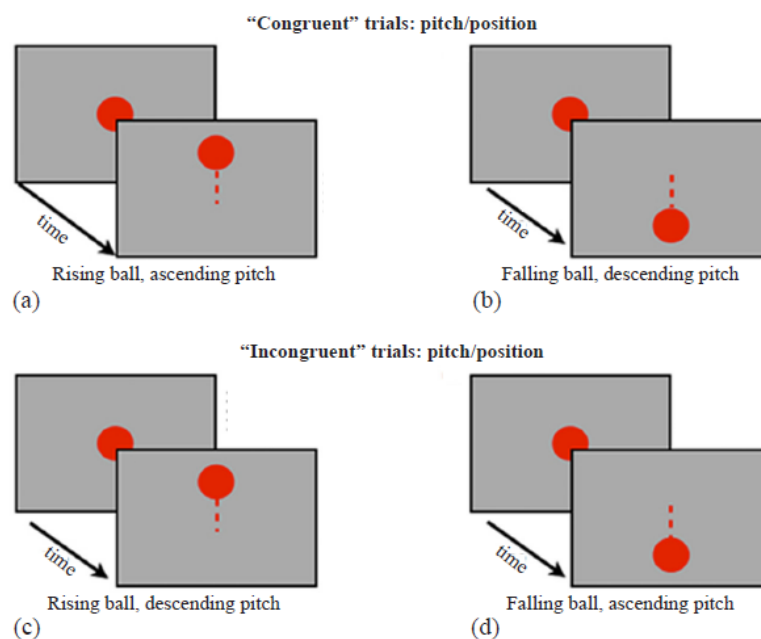


Figura 7. Ejemplo de secuencia de eventos para los estudios de Dolscheid, Hunnius, Casasanto y Majid (2014) y Walker y cols. (2010). (a) Condición congruente con frecuencia ascendente (b) Condición congruente con frecuencia descendente (c) Condición incongruente con frecuencia descendente (d) Condición incongruente con frecuencia ascendente (Figura adaptada de Parkinson, Kohler, Sievers, y Wheatley, 2012).

Las correspondencias transmodales también se han observado en niños pequeños (por ejemplo, 2,5 a 3 años de edad). En un estudio llevado a cabo por Mondloch y Maurer (2004), se pidió a los niños pequeños indicar cuál de dos pelotas que aparecían en la pantalla botando estaba produciendo un sonido grave o agudo, las pelotas variaban gradualmente, incrementando o decreciendo en su tamaño. Los resultados mostraron que los niños pequeños tendían a asociar el tono más alto y más bajo con un estímulo visual más pequeño y más grande, respectivamente. Obviamente, los niños de esta edad han estado ampliamente expuestos a las regularidades estadísticas del entorno y, en consecuencia, sus mecanismos de percepción y de asociación conceptual y semántica pueden haberse visto influidos por la exposición a dichas asociaciones. Por ejemplo, existen diversas asociaciones transmodales relevantes en el entorno del recién nacido, donde los objetos de menor tamaño tienden generalmente a producir sonidos más agudos que los objetos de mayor tamaño.

En la actualidad, no tenemos pruebas de que los niños nazcan con asociaciones transmodales. Además, si tenemos en cuenta que los bebés de 4 meses de edad (por tanto, prelingüísticos) ya han estado expuestos a una gran cantidad de estimulación multisensorial en su entorno, no existe evidencia que apoye o refute la hipótesis de que las correspondencias transmodales son innatas.

1.2.5 Bases neurales de la recodificación espacial de la altura tonal

La corteza auditiva primaria se organiza tonotópicamente (veáse Figura 9), es decir, cada frecuencia auditiva parece procesarse en un determinada área, implicando que las frecuencias más bajas están representadas en regiones laterales de la circunvolución de Heschl (también conocida como giro temporal transversal), mientras que frecuencias más altas estarían localizadas en regiones mediales de la circunvalación de Heschl (Humphries, Liebenthal, y Binder, 2010; Peretz y Zatorre, 2003) (ver Figura 8). Los daños en la circunvolución del Heschl pueden producir dificultad para percibir los cambios de frecuencia acústica (Zatorre, 1988; Johnsrude, Penhune, y Zatorre, 2000; Tramo, Shah, y Braidá, 2002).

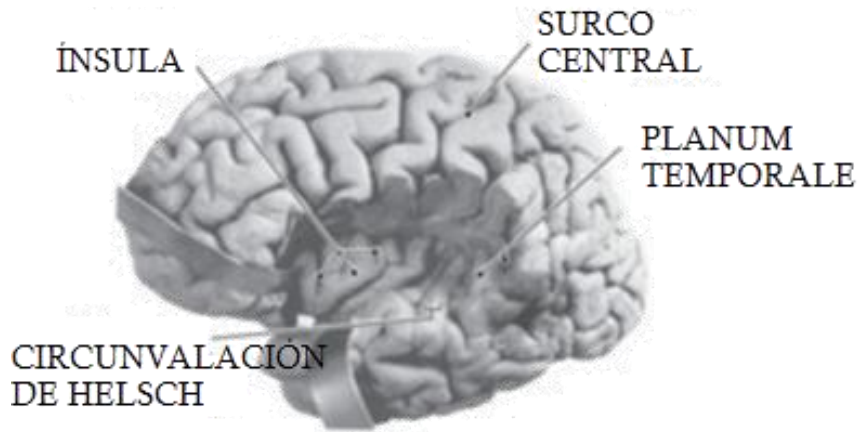


Figura 8. Circunvalación de Heschl ubicada en la corteza auditiva primaria.

El procesamiento de la altura tonal se atribuye generalmente al hemisferio derecho (HD) (Zatorre, 1984). Los pacientes con daño en el HD muestran una percepción acústica de la altura tonal más pobre que los pacientes con daño en el hemisferio izquierdo (HI) (Peretz, 1990; Sidtis, 1985). Además, la investigación en lesiones en el HD ha puesto de manifiesto la pérdida de habilidades en tareas de discriminación melódica o tonal (Peretz y Zatorre 2005). De hecho, el neocortex temporal derecho juega un papel esencial en el análisis de la frecuencia de los sonidos (Milner, 1962; Liégeois-Chauvel y cols., 1998; Zatorre, 1985).

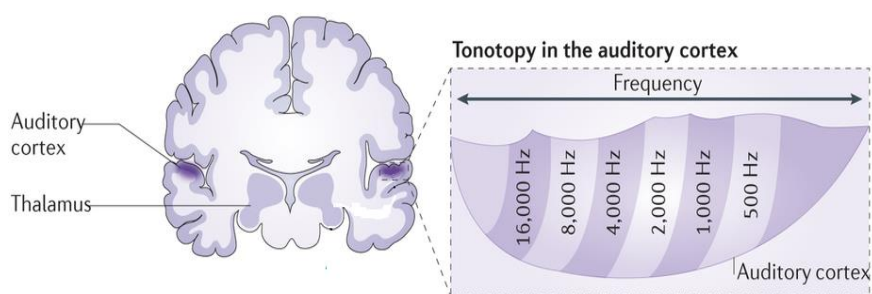


Figura 9. Representación tonotópica de la corteza auditiva primaria (Elgoyhen, Langguth, De Ridder, y Vanneste, 2015).

La posible representación espacial de la altura tonal podría ser esencial en la ejecución y percepción de la música (por ejemplo, en la notación musical) y para comprender la entonación y prosodia del habla. Estudios recientes, utilizando

técnicas de neuroimagen, han observado la activación en algunas zonas parietales específicas durante la ejecución de ciertas tareas musicales (Foster y Zatorre, 2010a, 2010b). En un estudio de Homae, Watanabe, Nakano, Asakawa, y Taga (2006), realizado con bebés de 3 meses de edad, mostraron a través de la técnica de topografía óptica de infrarrojo cercano, una mayor activación en las regiones temporo-parietales derechas al percibir un habla prosódica normal, es decir, variaciones en el tono y la intensidad, en comparación con las regiones temporo-parietales izquierdas (ver Figura 10) (Arimitsu y cols., 2011).

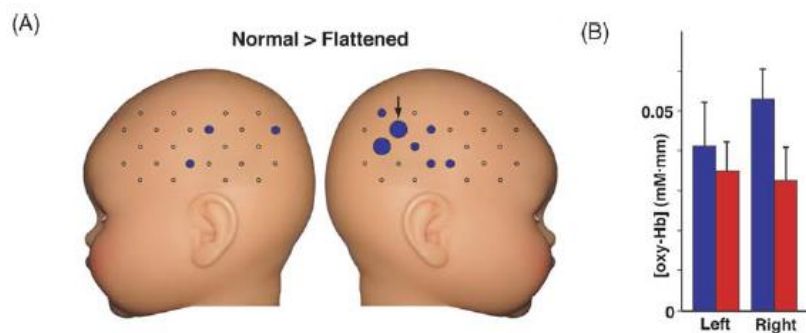


Figura 10. Resultados del estudio de Homae y cols. (A) Diferencias de activación entre el hemisferio izquierdo y derecho (B) Media de oxi-Hb. Las barras azules indican los cambios de señal bajo las condiciones de discurso con prosodia normal, y las barras rojas respecto a prosodia aplanada. Las barras de error indican el error estándar (Figura adaptada de Homae, Watanabe, Nakano, Asakawa, y Taga, 2006)

Las zonas parietales posteriores como el surco intraparietal inferior y superior (IPS) parecen estar involucradas en el análisis de la frecuencia del sonido (Foster y Zatorre, 2010a; 2010b) (ver Figura 11). Curiosamente, el IPS es una región multimodal que ha sido clásicamente relacionada el procesamiento espacial, por ejemplo, la coordinación perceptivo-motora, movimientos de agarre y manipulación de objetos, y operaciones aritméticas (Husain y Nachev, 2007; Kong y cols., 2005). Por tanto, el IPS realiza transformaciones sistemáticas sobre las representaciones sensoriales de estímulos visuoespaciales y táctiles (Grefkes y Fink, 2005).

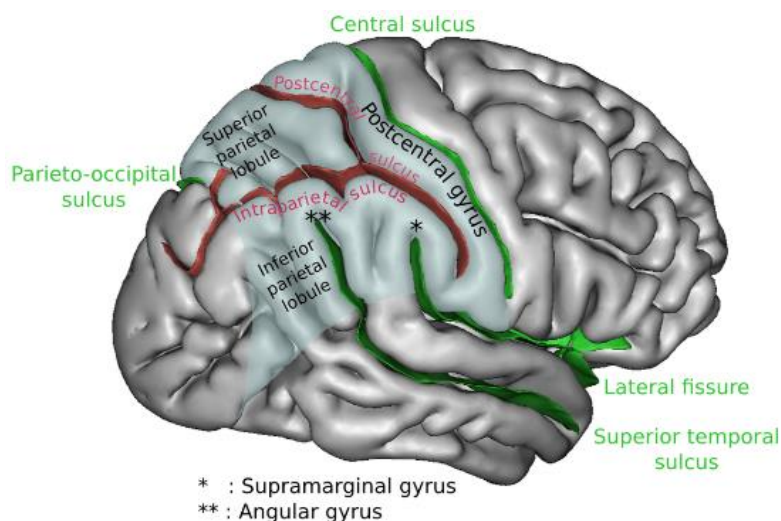


Figura 11. Representación de la superficie lateral del lóbulo parietal (Imagen reproducida con permiso: Sebastian 023, CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=20806983>)

Foster y Zatorre (2010a; 2010b) llevaron a cabo un estudio mediante la técnica de imagen por resonancia magnética funcional (IRMf) donde se reveló una activación del IPS en una tarea de transformación del patrón melódico. Participantes músicos y no músicos realizaron una serie de tareas auditivas: discriminación de melodía simple, discriminación de una melodía transpuesta, discriminación de ritmo melódico y discriminación fonémica (ver Figura 12). Todas las tareas siguieron el mismo procedimiento, los participantes tenían que juzgar si los patrones de dos secuencias de sonidos, musicales o fonémicos, eran iguales o diferentes respondiendo con el botón izquierdo o derecho de un ratón de ordenador (por ejemplo, si era la misma o diferente melodía).

Los resultados del estudio de Foster y Zatorre (2010a) mostraron activación funcional del IPS en la discriminación melodía transpuesta pero no en las otras tareas auditivas. A diferencia de la tarea de discriminación melódica simple, la tarea de transposición musical requería una transposición de 4 semitonos musicales, entre la primera y la segunda secuencia de sonidos, para detectar si existía variación entre los dos patrones. Los autores observaron un incremento significativo en la oxigenación cerebral de nivel-dependiente (*Brain oxygenation level-dependent*, BOLD) en el IPS anterior bilateral en la tarea de discriminación de melodía transpuesta, es decir, en

una operación de procesamiento de tono relativo (ver Figura 13). Las conclusiones de los autores fueron que el IPS juega un papel relevante en el procesamiento de la altura tonal (por ejemplo, la transposición musical), además de en las transformaciones visuoespaciales. Sin embargo, los autores no pudieron demostrar la actividad del IPS en tareas de discriminación de melodía simple. En consecuencia, parece que para evocar la activación del IPS en una tarea musical requiere de cierto nivel de procesamiento de transformación auditiva.

MELODY - ORIGINAL

RHYTHM - ORIGINAL

SIMPLE MELODY - DIFFERENT

RHYTHM - DIFFERENT

TRANPOSED MELODY - SAME

PHONEME PATTERN, ORIGINAL

KOH	ROO	RAH	NAH	FOO
-----	-----	-----	-----	-----

TRANPOSED MELODY - DIFF.

PHONEME PATTERN, DIFFERENT

KOH	ROO	RAH	YAH	FOO
-----	-----	-----	-----	-----

Figura 12. Estímulos de las cuatro tareas auditivas: melodía simple, melodía transpuesta, discriminación de ritmo y discriminación de fonemas. Flechas y líneas rojas indican alteración respecto al estímulo original (Foster y Zatorre, 2010a).

A raíz de la posible implicación de áreas posteriores del hemisferio derecho (por ejemplo, el IPS) en las tareas de tono acústico, no es descabellado pensar que un deterioro en habilidades visuoespaciales podría implicar una afectación la percepción de la frecuencia auditiva. Del mismo modo, una alteración de ciertas habilidades auditivas relacionadas con el procesamiento de la altura tonal podría indicarnos una posible afectación del procesamiento visuoespacial. Un ejemplo de esta doble afectación es la amusia, un síndrome neurológico en el cual los pacientes presentan déficits severos en el procesamiento de pequeños cambios tonales, así como en la memoria y el reconocimiento musical (Pearce, 2005).

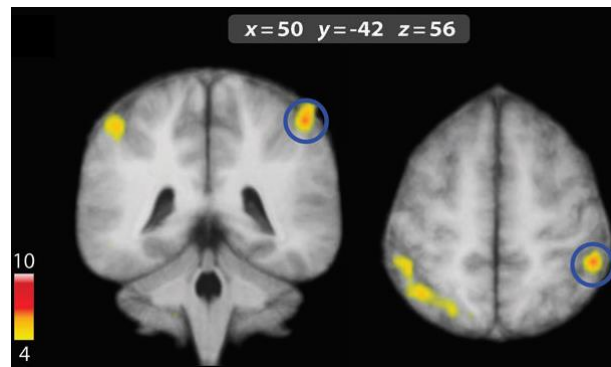


Figura 13. Señal BOLD en la tarea de transposición musical. El contraste de las tareas de discriminación de melodía simple y transpuesta revela una activación del IPS mayor en el HD.

La amusia se caracteriza principalmente por la dificultad en discriminar la frecuencia de los sonidos, hay una estimación de que este síndrome afecta a alrededor del 4% de la población (Peretz y Hyde, 2003; Sloboda, Wise, y Peretz, 2005). Douglas y Bilkey (2007) observaron que los pacientes con amusia tenían una ejecución significativamente peor en tareas de rotación mental, en comparación a un grupo control (Tillmann y cols., 2010, para resultados contradictorios). Douglas y Bilkey (2007) relacionaron estos resultados con el controvertido 'efecto Mozart'. El efecto Mozart podría inducir una mejoría a corto plazo en el desempeño de tareas espaciales (por ejemplo, rotación mental), debido a una breve exposición a determinadas piezas de música clásica (Rauscher, Shaw y Ky, 1993; Schellenberg, 2005).

Las conclusiones de los últimos estudios neuropsicológicos y de neuroimagen apoyan la hipótesis de que el procesamiento visuoespacial y el procesamiento de discriminación de frecuencias podrían compartir mecanismos neuronales.

2. OBJETIVOS

Diversos estudios y paradigmas muestran la existencia de correspondencias transmodales entre diversos atributos perceptivos (Spence, 2011).

Los tres objetivos de la presente disertación doctoral se centran en investigar las correspondencias transmodales entre varias dimensiones perceptivas específicas: la altura tonal, la elevación espacial y el tamaño. A continuación, se describen los objetivos específicos de la presente tesis:

Objetivo 1: Investigar si la detección de un estímulo visual en el espacio puede ser exógenamente facilitada por un determinado estímulo auditivo (por ejemplo, sonidos ascendentes) congruente presentado previamente. Para la consecución de este objetivo se realizó el Estudio 1 (Fernández-Prieto y Navarra, aceptado), donde se utilizó una modificación del paradigma atencional de Posner en participantes adultos sin experiencia musical (Posner, 1980). Los resultados del Estudio 1 demostraron que los sonidos ascendentes, pero no los descendentes, pueden modular la atención espacial en el eje vertical, pero no en el horizontal. Estos datos sugerirían que las frecuencias auditivas ascendentes pueden modular en mayor medida la atención visuoespacial en el eje vertical, en comparación a frecuencias descendentes.

Objetivo 2: Existen evidencias contradictorias respecto si los bebés prelingüísticos muestran asociaciones transmodales similares a los adultos, por ejemplo, entre la altura tonal y la localización espacial (Walker y cols., 2010; véase Lewkowicz y Minar, 2014 para resultados nulos). El objetivo fue responder a la cuestión de si la asociación audiovisual está presente en las primeras etapas de la vida y, por tanto, las asociaciones transmodales se desarrollan previamente a la adquisición de léxico espacial o abstracto (Swingley, 2009) o si es necesario una mayor exposición al medio y/o maduración de los procesos cognitivos implicados. Para responder a esta pregunta, se llevó a cabo el Estudio 2 (Fernández-Prieto, Navarra y Pons, 2015), donde se exploró la asociación entre la altura tonal y el tamaño de los estímulos visuales en bebés. Se testaron bebés de 4 y 6 meses de edad usando un paradigma de preferencia visual. Los resultados indicaron que los bebés de 6 meses si muestran

una asociación altura tonal-tamaño, pero no los bebés de 4 meses. Estos datos sugieren que es necesario un mínimo de experiencia y/o maduración cognitiva, y al mismo tiempo parece que el lenguaje no juega un papel esencial en dichas correspondencias transmodales.

Objetivo 3: Los resultados de los estudios 1 y 2 evidenciaron asociaciones transmodales en población sin alteraciones neurológicas, tanto en adultos como en bebés. En relación a estos resultados, el objetivo fue investigar si las alteraciones en zonas tradicionalmente asociadas con el procesamiento espacial (por ejemplo, áreas parietales del hemisferio derecho) podían producir déficits en la discriminación de la dirección de la altura tonal. Para cumplir este objetivo, realizamos el Estudio 3 (Fernández y cols., 2016). Se seleccionó una muestra de pacientes con trastorno del aprendizaje no verbal (TANV) y un grupo control. El TANV es un trastorno neurológico relacionado con alteraciones en el hemisferio derecho (Semrud-Clikeman y Hynd, 1990). Debido a que el TANV implica déficits en habilidades visuoespaciales, pudimos observar a través de diversas tareas visuales y auditivas si un grupo de pacientes preadolescentes y adolescentes con TANV también presentan anomalías discriminando altura tonal. Los resultados confirmaron que, además de déficits en las habilidades espaciales, el TANV también conlleva un déficit en la discriminación de variaciones en la altura tonal.

3. STUDY 1:

Fernández-Prieto, I. & Navarra, J. Spatial encoding of pitch in frequency sweeps (accepted, pending minor changes). *Psychology of Music*

Abstract

High-pitched sounds generate larger neural responses than low-pitched sounds. We investigated whether this difference influences the vertical representation of pitch. Participants performed a speeded detection of visual targets that could appear at one of four different spatial positions. Rising or falling frequency sweeps were randomly presented before the onset of the visual target. Faster reaction times to visual targets appearing above (but not below) a central fixation point were observed after the presentation of rising frequency sweeps. No significant effects were found for falling frequency sweeps and visual targets presented below fixation point. These results suggest that the difference in the level of arousal between rising and falling frequency influence their capacity of generating spatial representations. The fact that no difference was found, in terms of crossmodal effects, between the two upper positions may indicate that this 'spatial representation of pitch' is not specific for any particular spatial location but rather has a widespread influence over stimuli appearing in the upper visual field. The present findings are relevant for the study of music performance, human-machine interaction, the design of musical instruments, and research in other fields where visual and auditory stimuli with certain complexity are combined (music in advertisements, movies, etc.).

Introduction

Auditory pitch seems to be mapped into vertical coordinates in certain conditions (see Occelli, Spence, & Zampini, 2009; Rusconi, Kwan, Giordano, Umiltà, & Butterworth, 2006; Sonnadara, Gonzalez, Hansen, Elliott, & Lyons, 2009). Higher frequencies, for example, appear to be perceived as originating from higher positions in space (Bregman & Steiger, 1980; Pratt, 1930). Since the ancient Greeks, or even before, music notation has followed similar pitch-to-space mappings. For instance, the discovery of the Seikilos epitaph in Aydin (Turkey), from around AD 100 AD, gave us an example of how the visual representation of pitch in music notation followed this "crossmodal congruency" even in the ancient world.

The perceptual overlap between pitch and spatial elevation has widely been reported in the literature. Melara and O'Brien (1987) demonstrated that these two sensory dimensions are integrated in such a way that any variation perceived in one of them has a direct impact on a classification task involving the other. In this previous study, participants had to rapidly classify, by means of two different response buttons, two different tones (high vs. low) according to their pitch. Therefore, pitch was the task-relevant perceptual dimension. Each tone was presented together with a dot that could appear either above or below the visual middle line. The position of this visual stimulus was irrelevant for the task. The authors observed that participants' reaction times (RTs) were faster when the position of the dot was crossmodally congruent with the tone (e.g., the dot appeared above the middle line and was presented together with the high tone) than when they were incongruent. Moreover, the responses when classifying stimuli according to the relevant dimension slowed down as a consequence of variation perceived in the irrelevant dimension (Garner, 1974). Similar results were found when the participants had to judge the position of the dot and ignore any change in pitch.

Recent studies have also shown that reaction times (RTs) at judging differences in pitch between auditory stimuli can be modulated by the spatial location of the response button. The response to a sound that is higher in frequency with respect to a reference sound is faster when it implies an upward movement (Rusconi et al., 2006; see also Sonnadara et al., 2009). Similar results have been obtained using a large variety of experimental methods (see Occelli et al., 2009; Parise & Spence, 2009; Sonnadara et al., 2009), including indirect tasks (Lidji, Kolinsky, Lochy, & Morais, 2007; Rusconi et al., 2006). Further studies also suggest that the spatial representation of pitch can even modulate visuospatial attention (see Chiou & Rich, 2012; Mossbridge, Grabowecky, & Suzuki, 2011).

The fact that infants prefer to look at a visual stimulus that moves coherently with respect to a tone that progressively increases or decreases in frequency (Walker et al., 2012) suggests the presence of these perceptual association between auditory pitch and spatial elevation from the very first steps of life. In Walker and collaborators' study (2010), 3- to 4-months-old infants looked longer at visual stimuli that moved towards the upper part of the screen when they were presented together with a sound containing an ascending frequency sweep than when they were presented with a

sound with descending frequency. Dolscheid, Hunnius, Casasanto, and Majid (2012) followed the same procedure as Walker (2010) and found similar results showing crossmodal correspondences in prelinguistic babies. In contrast, however, Lewkowicz and Minar (2014) failed to replicate these effects after conducting five experiments with 4-, 6-, and 8-month-old infants, using both identical and different methods with respect to Walker et al. (2012; see also Walker et al. 2014, for a response to Lewkowicz & Minar's study).

Perceptual and cognitive differences between rising and falling pitch

Why is the music in climax scenes (e.g., in terror movies) so high-pitched? High frequencies are often perceived as being louder and more salient than low frequencies when they are presented at the same physical intensity (see Fletcher & Munson, 1933). In classic studies by Deutsch (1976, 1978), participants listened to two different sounds played simultaneously, each one presented at a different ear, with frequencies of 400 and 800Hz. They reported to hear a fused tone at the ear where the higher tone was presented. This phenomenon could be taken as evidence suggesting that the higher frequencies have a larger influence on the spatial perception (i.e., lateralization) of the stimuli than the lower frequencies (see also Deutsch & Roll, 1976; von Békésy, 1963).

Several studies conducted with infants indicate a preference for high frequencies at early stages of life. Infants tend to show predilection for listening to high-pitched than low-pitched speech (Patterson et. al 1997) and songs (Trainor & Zacharias, 1998). Furthermore, the discrimination between high frequencies seems to precede, during the maturation of infants' auditory system, the discrimination between low frequencies (Olsho, 1984; Olsho, Koch, & Halpin, 1987; Trehub, Schneider, & Endman, 1980). At a neural level, several studies have revealed, using electroencephalography (EEG), larger mismatch negativity (MMN) for higher than for lower deviant tones (Näätänen, 1990; Naatanen, Gaillard, & Mantysalo, 1978; Ruusuvirta & Astikainen, 2012). Previous research using tones that contained frequency sweeps (i.e., dynamically covering a range of frequencies) has revealed a better performance at detecting an increase in frequency than a decrease (Kishon-

Rabin, Roth, Dijk, Yinon, & Amir, 2004). In the so-called Doppler Shift, an increase in acoustic frequency is perceived for sounds that are associated to visual stimuli that approach us, even when these sounds are presented at a constant frequency (see Neuhoff, McBeath, & Wanzie, 1996; see also Hasset, 1999; McBeath & Neuhoff, 2002). This effect suggests the presence of a phenomenological relation between the perception of rising pitch and the alertness generated by objects that move towards us. This illusory percept has been related to the *Doppler Effect* (Doppler, 1842), in which the frequency of a sound generated by moving object is perceived to be higher, identical and lower than the emitted frequency as the object approaches, passes by and recedes an observer, respectively. For example, the perceived sound of an ambulance siren is perceived as being higher or lower in frequency than the emitted frequency when the ambulance approaches or moves away from the perceiver, respectively. Thus, there seems to be an association between rising frequencies, approaching objects and, arguably, an increase of alertness. In contrast, dynamic sounds with descending frequencies seem to be more related to objects that move away, perhaps inducing a reduction of the level of alertness. This evidence could easily lead to the hypothesis that rising (i.e., low-to-high) frequency sweeps have a larger impact on arousal and/or alertness than descending (i.e., high-to-low) frequency sweeps. An interesting question, addressed in the present study, refers to the possibility that ascending frequency sweeps (or high tones) also have larger inherent spatial properties than descending frequency sweeps (or low tones). This has not been directly explored in previous studies, where possible crossmodal effects have been addressed including collapsed data from both high (or ascending) and low (or descending) pitch (e.g., Rusconi et al., 2006; Sonnadara et al., 2009.).

Due to the physical properties of the sound and the spatial separation between the ears' auricles, humans and other animals can localize high frequencies more accurately than low frequencies (see Masterton, Heffner, & Ravizza, 1969). Unlike low frequencies, which are characterized by having long wavelengths, high frequencies have short wavelengths that allow sound localization based on interaural time difference (i.e., the interval between the different moments at which an acoustic signal reaches each ear). In consequence, the spatial localization of sound relies more on high frequencies than on low frequencies. This fact, combined with the idea that high frequency sounds generate a larger physiological response, could make us expect

larger crossmodal effects in sounds with rising pitch than in sounds with descending pitch.

In the present study, we investigated whether tones with rising frequencies are better suited for generating spatial representations than tones with falling frequencies. For this purpose, an adaptation of Posner's cueing paradigm (Posner, 1980) was used. This paradigm was originally designed to study spatial attention. Two boxes were presented on the left and the right of the screen. One of these two boxes was briefly highlighted (spatial cue). After a certain stimulus-onset asynchrony (SOA), an asterisk (visual target) appeared in the same box as the spatial cue (i.e., valid or congruent trial) or else in the box at the opposite side (i.e., invalid or incongruent trial). Participants were faster at detecting the asterisk in congruent trials than in incongruent trials. Therefore, the detection of a visual target could be facilitated by a previous visuospatial cue that oriented spatial attention towards a specific area of the visual field.

In our modification of Posner's paradigm, rising and falling frequency sweeps were used as spatial cues, under the assumption that they may differ in their capability to modulate (1) the perceiver's arousal and alertness (see Tomatis, 1978), and consequently, (2) visuospatial attention. More specifically, we used Posner's cueing paradigm to test the hypothesis that the detection of visual targets would be more affected by rising frequency sweeps than by falling frequency sweeps. While the majority of studies addressing the spatial representation of pitch have used pure tones, presented either in isolation or else embedded in melodies, only a few of them used pitch-varying (dynamic) stimuli such as frequency sweeps (see Mossbridge, Grabowecky, & Suzuki, 2011; Walker et al., 2010 for exceptions). Because of the possible capability of frequency dynamic sweeps to induce "spatial directionality", they are particularly useful for the study of the spatial representation of pitch.

Vertical and horizontal crossmodal effects between pitch and spatial elevation

While the vertical representation of pitch has received most of the attention in research on crossmodal correspondences, only a few studies have addressed the

possible perceptual correspondence between pitch and space in the horizontal axis (see Rusconi et al., 2006; Lidji et al., 2007). The possible horizontal representation of pitch could perhaps be related to the ability of the brain to create a mental representation of quantities (e.g., with small numbers being located on the left side and the large numbers on the right side; see Dehaene, Bossini, & Giraux, 1993; see also Hubbard, Piazza, Pinel, & Dehaene, 2005, for a review). Following an auditory-based reinterpretation of this metaphorical association, low frequencies would be located on the left side (in Western cultures, at least; see Dehaene, Bossini, & Giraux, 1993) and high frequencies on the right. This possible crossmodal association may perhaps explain why musical instruments are often designed following this "left-low/right-high rule" (e.g., piano).

Rusconi and colleagues (2006) reported crossmodal correspondence effects between pitch and horizontal space in experienced musicians. In one of the experiments included in their study, participants completed a speeded pitch discrimination task where they had to compare the frequency of a probe and a reference sound. The results showed that the response times (RTs) were modulated by the spatial location of the response button in the horizontal axis: faster reaction times were observed for "higher" and "lower" responses when participants had to press a key located at the right and at the left side of the keyboard, respectively. However, the fact that no effects were observed in non-musicians that also participated in this study suggests that horizontal representations of pitch are largely driven by experience (e.g., musical training; see also Mossbridge et al., 2011; Chiou & Rich, 2012). According to this speculative interpretation, pitch would preferentially be represented vertically. The experimental approach adopted in the present study (see Figure 1) allowed us to investigate this possible vertical-over-horizontal preference in the spatial representation of pitch in non-musicians.

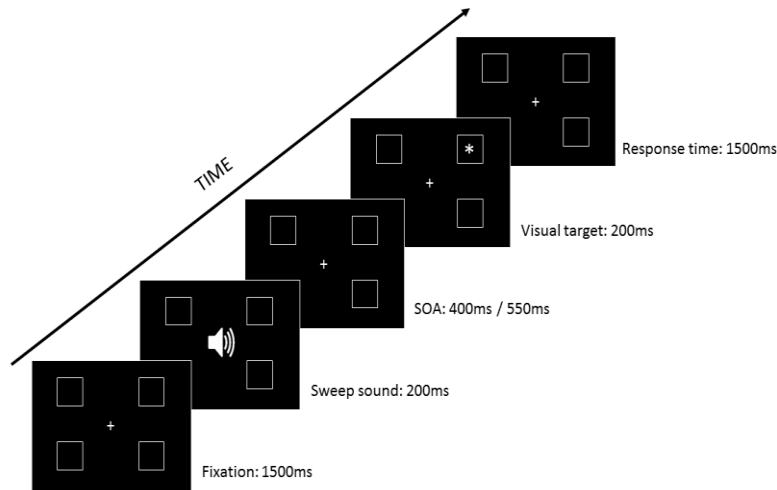


Figure 1. Sequence and timing of the events in a trial of the experiment. One of the two different frequency sweeps (one varying from 200Hz to 700Hz and another from 700Hz to 200Hz) were presented in an adapted version of the classical Posner paradigm.

How specific is the vertical representation of pitch?

Using an adaptation of the Posner cuing paradigm (Posner, 1980), Mossbridge and colleagues (2012) recently showed that perceiving low-to-high (rising) and high-to-low (falling) frequency sweeps facilitated the subsequent detection of a visual stimulus appearing in a spatial position (upper or lower) that was crossmodally congruent with the sound. Interestingly, this crossmodal correspondence effect vanished when the visual stimulus appeared in one of 4 different positions (left-up corner, right-up corner, left-down corner or right-down corner of the computer screen; see Figure 2), instead of centrally (i.e., above or below a central fixation point). This pattern of results supports a "local", rather than "global", account of the spatial representation of pitch, as sounds only influenced the processing of visual stimuli that appeared at a relatively small area immediately above or below the gaze's fixation point.

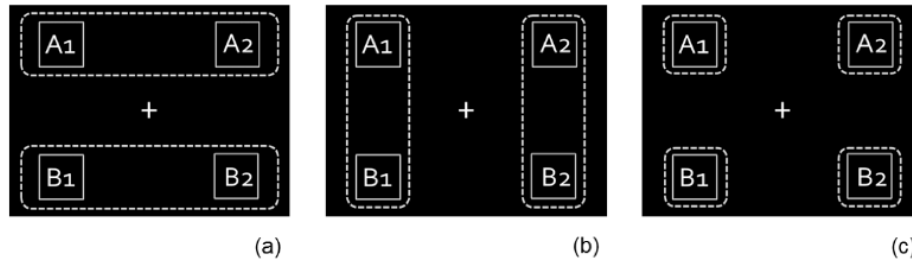


Figure 2. Global Vs. Positional Analysis. (a) Global vertical cueing analysis. Targets appearing upper positions of the screen (A1 and A2) and lower positions of the screen (B1 and B2) (b) Global horizontal cueing analysis. Targets appearing on the left (A1 and B1) and right (A2 and B2) sides of the screen. (c) Positional cueing analysis. Targets appearing on specific spatial positions (A1, A2, B1 and B2).

Another relevant aspect of Mossbridge et al.'s study (2012) is that the frequency sweeps only ranged from 300 to 450Hz. If pitch can effectively be mapped into spatial coordinates, a possibility may be that a larger variation (e.g., 500Hz instead of just 150Hz) could 'cover' a larger area of the visual space, thus inducing larger cuing effects. Indeed, a plausible hypothesis, tested in the present study, could be that frequency sweeps covering a larger range of sound frequencies (e.g. 200 to 700Hz or 700 to 200Hz) generate a "path" between 2 relatively specific spatial positions (e.g., between positions B1 and A2 in; see Figure 2).

Keeping in mind that that the optimal vertical remapping sound seems to take a certain amount of processing time (i.e., more than 300ms; see Chiou et al., 2012), two relatively long stimulus onset asynchronies SOAs (400ms and 550ms) were also used, between the tone and the visual stimulus, to allow for a complete spatial remapping of sound, thus facilitating the appearance of spatial cuing effects.

Methods

Participants

Sixteen right-handed non-musician participants (12 females, average age: 21.5), with normal hearing, and normal or corrected-to-normal vision, took part in the study, and received 6 euros for their participation. None of the participants received musical

training since elementary school. The experiment was conducted in accordance with the Declaration of Helsinki, and had ethical approval from the XXX Ethics Committee (information omitted to keep anonymity). The participants provided a written informed consent to participate in the study.

Materials

An Intel Core computer and a 15-inch CRT monitor (Philips 107-E Monitor, 85 Hz) were used for testing. The experimental procedure was run by E-Prime 2.0 (Psychology Software Tools Inc., Pittsburg, PA) in a dark and soundproof room. The participants sat at a table in front of the monitor at an approximate distance of 60cm. Two loudspeakers (Phillips A 1.2 Fun Power, 7510704863, China) were located at each side of the computer screen.

Procedure

In the variation of the Posner's cueing paradigm (Posner, 1980) employed in the present study, the participant had to detect a visual target (a white asterisk of 1.3 cm of diameter) as fast and as accurately as possible in 320 trials. The visual target could appear at one of four different spatial positions (two above and two below the fixation point; see Figure 2). Rising (200-700Hz) or falling (700-200Hz) frequency sweeps were randomly presented either 400 or 550 ms before the onset of a visual target. These two different SOAs were selected based on previous literature (Chiou & Rich, 2012) and were randomly presented to avoid temporal predictability between the auditory and the visual stimuli. Participants were instructed to press, as fast as possible, using the index finger of their right hand, a key on a computer keyboard after detecting a visual target. The participants' index finger of the right hand rested on the response key during the testing session.

Each trial began with a fixation display (1.5x1.5 cm), consisting of a white central cross flanked by four square-shaped placeholders of 16 cm² (two above and two below the fixation point; see Figure 2). After 1500 ms, one of the two possible frequency sweeps (rising or falling; constantly varying from 200 to 700Hz or from 700 to 200Hz, respectively) was presented for 210 ms (with a 5-ms fade-in and fade-out to avoid clicks) at 75dB(A). The visual target appeared, for 200 ms, after a stimulus-onset

asynchrony (SOA) of either 400 or 550 ms, at one of four different spatial positions inside a placeholder (up-right, up-left, down-right or down-left; i.e., positions A1, A2, B1, and B2 in Figure 2). The display (fixation cross and placeholders) remained visible until participants' response with a time limit of 1500 ms.

Results

Reaction times (RTs) faster than 150 ms were considered as anticipatory responses and were not included in the statistical analyses. This decision was motivated by the fact that the time needed to process the perceptual information and execute the motor response cannot physiologically be shorter than 150ms. Several statistical analyses were performed to address possible audiovisual cuing effects along the vertical axis (A1+A2 vs. B1+B2; see Figure 2), along the horizontal axis (A1+B1 vs. A2+B2; see Figure 2), as well as at specific positions on each of the different placeholders (A1 vs. A2 vs. B1 vs. B2; see Figure 2).

'Vertical' analyses

Following previous literature (see Spence, 2011, for a review), we understood the factor "spatial congruence" as follows: A target that appeared in an upper position of the screen (A1 or A2 positions) after the presentation of a rising frequency sweep was considered as congruent. Visual targets appearing at lower positions (B1 or B2) after a rising frequency sweeps were considered as incongruent. In contrast, visual targets appearing at a lower position (B1 or B2) or at a higher position (A1 or A2), after the presentation of a falling frequency sweep, were considered as congruent and incongruent, respectively.

A repeated-measures analysis of variance (ANOVA), including the within-subjects factors "congruency" (congruent vs. incongruent condition), "SOA" (400 vs. 550ms) and "direction of the frequency sweep" (ascending vs. descending frequency sweep), revealed no interaction between this three factors, $F(1, 15) = .835$, $p = .375$, $\eta_p^2 = .053$. Significant effects of congruency and direction of the frequency sweep were found, $F(1, 15) = 5.996$, $p = .027$, $\eta_p^2 = .236$; $F(1, 15) = 5.325$, $p = .036$, $\eta_p^2 = .262$,

respectively. A subsequent ANOVA including only the factors “congruency” and “direction of the frequency sweep” revealed a significant interaction between them, $F(1, 15) = 6.695, p = .018, \eta_p^2 = .318$. No significant differences were found between the two SOAs.

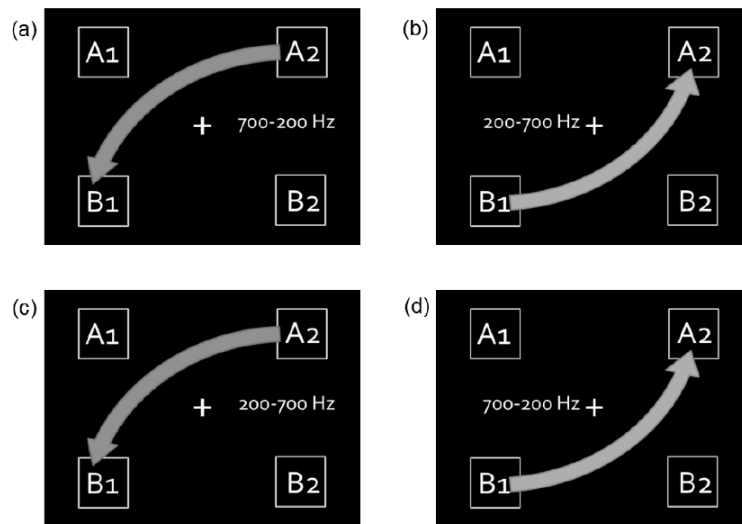


Figure 3. The arrow represents a hypothetical dynamic spatial representation of the auditory cue. (a) Congruent condition (valid cue) with a frequency sweep sound going from a high tone to low tone (b) Congruent condition (valid cue) with a frequency sweep sound going from a low tone to high tone (c) Incongruent condition (invalid cue) with a frequency sweep sound going from a high tone to low tone (d) Incongruent condition (invalid cue) with a sweeper sound going from a low tone to high tone.

Further analyses, including collapsed data from both SOAs and conducted only with trials that contained rising frequency sweeps, revealed significantly faster RTs in the congruent condition than in the incongruent condition ($t(15) = -4.513, p < .01$). No significant differences were found between the congruent and the incongruent condition for falling frequency sweeps ($t(15) = .230, p = .822$) (see Table 1 and Figure 4).

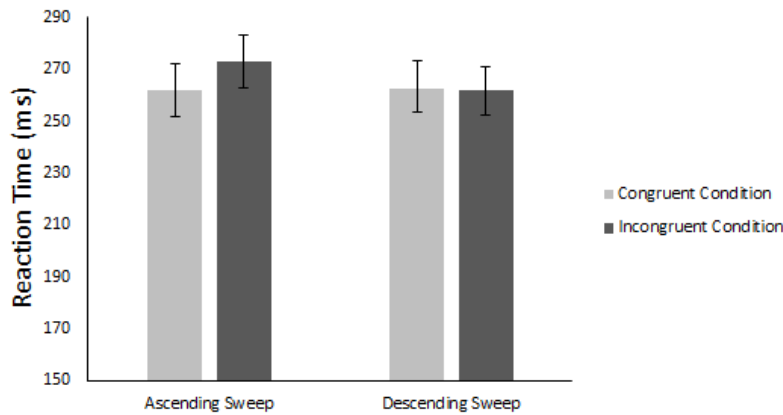


Figure 4. Mean of the participant (in milliseconds) in each condition (congruent/incongruent) for the two different frequency sweeps (ascending/descending). Error bars indicate standard error of the mean.

'Horizontal' analyses

In the "horizontal" analyses, and following previous literature (see Rusconi et al., 2006), targets appearing on the left (A1 and B1) and right (A2 and B2) sides of the screen were considered as congruent after the presentation of falling and rising frequency sweeps, respectively (see Figure 3). Targets appearing on the opposite sides were considered as incongruent. None of the significant effects found in the previous analyses were observed in the 'horizontal' analyses (see Table 1).

'Positional' analyses

An ANOVA including the within-subjects factors "SOA", "position" (A1, A2, B1, B2; see Figure 3) and "direction of the frequency sweep" revealed an absence of interaction between the 3 factors, $F(1, 15) = 1.455, p = .272, \eta_p^2 = .251$. However, significant differences by position were revealed, $F(1, 15) = 6.134, p = .008, \eta_p^2 = .686$; reflecting the widely-observed tendency of upward movements to be faster than downward movements.

In an attempt to see whether the spatial cuing effects were "global" (i.e., taking place on the upper or lower wide areas of the screen, including A1+A2 and B1+B2, respectively) or "local" (i.e., for specific spatial positions, e.g., A2), more analyses were carried out. T-tests, conducted separately for rising and falling frequency

sweeps, revealed no significant differences between the RTs in each of the four different positions and the average of RTs in all of the other positions. These analyses allowed us to see whether RTs to visual targets were significantly faster or slower (see Table 1), when preceded by a specific auditory cue, in a particular position (e.g., A2; see Figure 3b) than in the other three positions (A1, B1, and B2). Note that the tests for a possible "global" account of our results (i.e., considering collapsed data from the two upper positions and the two lower positions separately) are presented above (see 'vertical' analyses).

Table 1. Participants' mean reaction times and standard deviations for all conditions. Student's *t*-tests between congruent and incongruent conditions.

	Direction of the frequency sweep	Congruent condition	Incongruent Condition	Student's <i>t</i> -test
Vertical Cueing effects	Rising pitch	261,917 ms (SD = 41,276 ms)	272,939 ms (SD = 43,992 ms)	$t(15) = -4.513, p < .01$
	Falling pitch	262,620 ms (SD = 40,408 ms)	261,800 ms (SD = 40,408 ms)	$t(15) = .230, p = .822$
Horizontal Cueing effects	Rising pitch	263,978 ms (SD = 42,010 ms)	261,356 ms (SD = 41,062 ms)	$t(15) = 1.314, p = .208$
	Falling pitch	263,375 ms (SD = 38,747 ms)	263,096 ms (SD = 43,382 ms)	$t(15) = -.657, p = .521$
Positional Cueing effects	Rising pitch	263,636 ms (SD = 40,490 ms)	268,692 ms (SD = 43,402 ms)	$t(15) = -1.684, p = .113$
	Falling pitch	264,668 ms (SD = 40,312 ms)	261,391 ms (SD = 37,899 ms)	$t(15) = 1.042, p = .314$

Discussion

Along with previous studies (Rusconi et al., 2006; Sonnadara et al., 2009; Chiou & Rich, 2012) our results suggest that auditory stimuli have inherent spatial properties that can modulate the subsequent spatial processing of visual stimuli by means of

spatial cuing. However, this cuing effect was only observed, in our study, in certain conditions:

- (1) Rising frequency sweeps elicited faster responses to visual targets presented on the superior part of the screen (see Figure 2) than to visual targets presented on the inferior part of the screen. Falling frequency sweeps did not elicit comparable effects (i.e., faster RTs for visual stimuli presented below fixation point). The fact that crossmodal correspondences occurred for rising but not for falling frequency sweeps may be related to basic differences between them in terms of modulating the perceivers' physiological response (Ruusurvirta & Astikainen, 2012; Näätänen et al., 1978; Näätänen, 1990). Following our initial hypotheses, this difference between rising and falling frequency sweeps may percolate into their capacity to generate spatial representations and cuing effects based on crossmodal correspondences
- (2) Frequency sweeps do not modulate, in non-musicians, the detection of visual target appearing at the right or the left side of the visual field (horizontal axis). Crossmodal correspondences between pitch and space occurred along the vertical axis but not along the horizontal axis, suggesting that pitch is preferentially represented vertically, rather than horizontally.
- (3) The spatial cueing effect generated by the rising frequency sweeps was, in our study, not specific for any particular position (e.g., right-up corner, or A2 in our experiment). This result supports a more 'global' (i.e., for 'up' and 'down' positions, in general) than 'local' (i.e., for a particular position in space) account of the spatial representation of pitch.
- (4) Rising frequency sweeps seemed to slow down the participants' responses to visual stimuli that appeared in a crossmodally incongruent spatial position (i.e., below fixation point). However, a neutral (or baseline) condition would be needed to further test this hypothesis and see whether RT effects can be seen in congruent trials (i.e., shorter RTs with respect to neutral trials), incongruent trials (i.e., slower RTs) or in both types of trial.

A possible explanation of why the cueing effects took place only for upper positions may be that low-to-high (rising) frequency sweeps have a larger impact over the perceiver's arousal than falling sweeps, and also that this effect interacts with the crossmodal (spatial) representation of pitch. In line with classical sound lateralization studies (Deutsch, 1976, 1978; Deutsch & Roll, 1976), the auditory system presents a perceptual bias for high frequencies over lower frequencies. Higher frequencies seem to drive sound localization: competitive sounds are usually perceived in the ear that received the highest frequency. We believe that crossmodal correspondences occur more preponderantly for ascending frequency sweeps due to the particular properties of high (and, arguably, ascending) frequencies: They can, for example, increase the perceiver's alertness (Tomatis, 1978). The fact that high-pitched sounds generate more psycho-physiological response (as measured in EEG) than low-pitched tones (see Näätänen, 1990; Naatanen, Gaillard, & Mantysalo, 1978; Ruusuvirta & Astikainen, 2012) may also support our interpretation of the results. At an early age, our auditory system seems to be more tuned to perceive high frequencies than low frequencies (Olsho, 1984; Olsho, Koch, & Halpin, 1987; Trehub, Schneider, & Endman, 1980).

Our data did not indicate the presence of any spatial representation of pitch along the horizontal axis in non-musicians. Despite the fact that intense musical training can modify and increase the spatial encoding along the horizontal axis (see Rusconi et al., 2006), our results indicate that pitch is predominantly encoded vertically in absence of intensive musical training. Furthermore, and considering evidence from a previous study by Lidji's and collaborators (2007), the use of indirect speeded tasks (e.g., detecting a visual target, as in the present study) does not seem to produce crossmodal correspondence effects between pitch and space in the horizontal plane in non-musicians. Further results obtained by Chiou & Rich (2012), in which no evidence was found indicating the presence of crossmodal correspondence along the horizontal axis in non-musicians, may provide further support for the idea that pitch is preferentially represented vertically in listeners with no musical expertise.

As Figure 3 reveals, the frequency sweep used in the present study could plausibly have originated a "sense of direction", perhaps moving the focus of attention to a specific area of the superior or inferior visual field (e.g., a rising frequency sweep cueing

position A2). However, our data suggests that the effects of the spatial representation of rising pitch are widespread and non-directional rather than location-specific and directional.

Finally, the larger RTs observed in the incongruent condition are in line with another recent study conducted in our laboratory, in which a larger amplitude of the P3b visual-evoked potential was observed as consequence of a mismatch between a “spatial expectation” generated by a highly-predictable melody and the spatial location of a visual target (XXX, submitted, information omitted to information omitted to keep anonymity). Taken together, the results of both studies may suggest that auditory stimuli containing changes in pitch (e.g., sounds with rising pitch) can modulate the perceptual system's reaction to upcoming visual targets that appear in particular spatial positions. Thus, these findings could perhaps have a significant impact in several disciplines related to sound processing and music. The design of musical instruments, loudspeakers, or digital platforms for music production and edition may perhaps take our results into account to balance (or take advantage of) the different psychological effects of perceiving low and high frequencies.

Funding

The present study was supported by a fellowship from Institut de Recerca en Cervell, Cognition & Behaviour (IR3C; Universitat de Barcelona) & Fundació Sant Joan de Déu to I.F.-P. and by grant PSI2012-39149 from Ministerio de Economía y Competitividad (Spanish Government) to J.N.

References

- Bendor, D., & Wang, X. (2006). Cortical representations of pitch in monkeys and humans. *Current opinion in neurobiology*, 16(4), 391-399.
- Bregman, A., & Steiger, H. (1980). Auditory streaming and vertical localization: Interdependence of “what” and “where” decisions in audition. *Percept Psychophys*, 28(6), 539-546
- Chiou, R., & Rich, A. N. (2012). Cross-modality correspondence between pitch and spatial location modulates attentional orienting. *Perception*, 41(3), 339-353.
- Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and numerical magnitude. *J. Exp. Psychol. Gen.*, 122, 371-396.
- Deutsch, D. (1976). Lateralization by frequency in dichotic tonal sequences as a function of interaural amplitude and time differences. *J Acoust Soc Am*, 60(S1), S50-S50.
- Deutsch, D. (1978). Lateralization by frequency for repeating sequences of dichotic 400- and 800-Hz tones. *J Acoust Soc Am*, 63(1), 184-186.
- Deutsch, D., & Roll, P. L. (1976). Separate "what" and "where" decision mechanisms in processing a dichotic tonal sequence. *J Exp Psychol Hum Percept Perform*, 2(1), 23-29.
- Dolscheid, S., Hunnius, S., Casasanto, D., & Majid, A. (2014). Prelinguistic Infants Are Sensitive to Space-Pitch Associations Found Across Cultures. *Psychol Sci*, 25(6), 1256-1261.
- Doppler, C. (1842). Ueber das farbige Licht der Doppelsterne und einiger anderer Gestirne des Himmels: Versuch einer das Bradley'sche Aberrations-Theorem als integrirenden Theil in sich schliessenden allgemeineren Theorie. Borrosch & André.
- Fletcher, H., & Munson, W. A. (1933). Loudness, Its Definition, Measurement and Calculation. *The Journal of the Acoustical Society of America*, 5(2), 82-108.
- Garner, W. R. (1974). *The processing of information and structure*. Potomac, MD: Lawrence Erlbaum Associates, Inc.
- Hassett, S. M., & Feth, L. L. (1999). Just discriminable change of velocity of a simulated moving sound source. *The Journal of the Acoustical Society of America*, 106(4), 2209-2209.

- Hubbard, E. M., Piazza, M., Pinel, P., & Dehaene, S. (2005). Interactions between number and space in parietal cortex. *Nat Rev Neurosci*, *6*(6), 435-448
- Kishon-Rabin, L., Roth, D. A. E., Dijk, B. V., Yinon, T., & Amir, O. (2004). Frequency discrimination thresholds: the effect of increment versus decrement detection of frequency. *Journal of basic and clinical physiology and pharmacology*, *15*(1-2), 29-40.
- Lewkowicz, D. J., & Minar, N. J. (2014). Infants Are Not Sensitive to Synesthetic Cross-Modality Correspondences: A Comment on Walker et al. (2010). *Psychol Sci*.
- Lidji, P., Kolinsky, R., Lochy, A., & Morais, J. (2007). Spatial associations for musical stimuli: A piano in the head? *Journal of Experimental Psychology-Human Perception and Performance*, *33*(5), 1189-1207.
- Masterton, B., Heffner, H., & Ravizza, R. (1969). The evolution of human hearing. *The Journal of the Acoustical Society of America*, *45*(4), 966-985.
- Mcbeath, M. K., & Neuhoff, J. G. (2002). The Doppler Effect is not what you think it is: Dramatic pitch change due to dynamic intensity change. *Psychonomic bulletin & review*, *9*(2), 306-313.
- Melara, R. D., & O'Brien, T. P. (1987). Interaction between synesthetically corresponding dimensions. *Journal of Experimental Psychology: General*, *116*(4), 323-336.
- Morton, E. S. (1977). On the occurrence and significance of motivation structural rules in some bird and mammal sounds. *American Naturalist*, *111*, 855-869.
- Mossbridge, J. A., Grabowecky, M., & Suzuki, S. (2011). Changes in auditory frequency guide visual-spatial attention. *Cognition*, *121*(1), 133-139.
- Näätänen, R. (1990). The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behavioral and Brain Sciences*, *13*(02), 201-233.
- Näätänen, R., Gaillard, A. W., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta psychologica*, *42*(4), 313-329.
- Neuhoff, J. G., McBeath, M. K., & Wanzie, W. C. (1999). Dynamic frequency change influences loudness perception: a central, analytic process. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(4), 1050.

- Occelli, V., Spence, C., & Zampini, M. (2009). Compatibility effects between sound frequency and tactile elevation. *Neuroreport*, *20*(8), 793-797.
- Olsho, L. W. (1984). Infant frequency discrimination. *Infant behavior and development*, *7*(1), 27-35.
- Olsho, L. W., Koch, E. G., & Halpin, C. F. (1987). Level and age effects in infant frequency discrimination. *J Acoust Soc Am*, *82*(2), 454-464.
- Parise, C. V., & Spence, C. (2009). When Birds of a Feather Flock Together: Synesthetic Correspondences Modulate Audiovisual Integration in Non-Synesthetes. *PLoS One*, *4*(5), e5664.
- Patterson, M., Muir, D. W., & Hains, S. (1997). *Infant sensitivity to perturbations in adult infant-directed speech during social interactions with mother and stranger*. Poster presented at the Society for Research in Child Development, New Orleans.
- Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*(1), 3-25.
- Pratt, C. C. (1930). The spatial character of high and low tones. *J. Exp. Psychol.*, *13*, 278-285.
- Rusconi, E., Kwan, B., Giordano, B. L., Umiltà, C., & Butterworth, B. (2006). Spatial representation of pitch height: the SMARC effect. *Cognition*, *99*(2), 113-129.
- Ruusuvirta, T. T., & Astikainen, P. (2012). Mismatch negativity of higher amplitude for melodic ascendance than descendance. *Neuroreport*, *23*(4), 220-223.
- Sonnadara, R. R., Gonzalez, D. A., Hansen, S., Elliott, D., & Lyons, J. L. (2009). Spatial properties of perceived pitch: influence on reaching movements. *Annals of the New York Academy of Sciences*, *1169*, 503-507.
- Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, *73*(4), 971-995.
- Tomatis, A.A. (1978) *Presented in International Kodaly Symposium in Paris*.
- Trainor, L. J., & Zacharias, C. A. (1998). Infants prefer higher-pitched singing. *Infant behavior and development*, *21*(4), 799-805.
- Trehub, S. E., Schneider, B. A., & Endman, M. (1980). Developmental changes in infants' sensitivity to octave-band noises. *Journal of Experimental Child Psychology*, *29*(2), 282-293

- von Békésy, G. (1963). Three Experiments Concerned with Pitch Perception. *J Acoust Soc Am*, 35(4), 602-606.
- Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A., & Johnson, S. P. (2010). Preverbal infants' sensitivity to synaesthetic cross-modality correspondences. *Psychol Sci*, 21(1), 21-25.
- Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A., & Johnson, S. P. (2014). Preverbal Infants Are Sensitive to Cross-Sensory Correspondences: Much Ado About the Null Results of Lewkowicz and Minar (2014). *Psychol Sci*, 25(3), 835-836.

4. STUDY 2:

Fernández-Prieto, I., Navarra, J. & Pons, F. How big is this sound?

(2015) Crossmodal association between pitch and size in infants.

Infant behavior and development, 38, 77-81.

<http://dx.doi.org/10.1016/j.infbeh.2014.12.008>

Abstract

High-frequency sounds are associated crossmodally with high spatial positions, as well as with bright, small, and sharp objects. Conversely, low frequency sounds are associated with low spatial positions, and with dark, big and rounded objects (see Bien, ten Oever, Goebel, & Sack, 2012; Gallace & Spence, 2006; Marks, 1974; Melara & O'Brien, 1987; Rusconi, Kwan, Giordano, Umiltà, & Butterworth, 2006; see also Spence, 2011, for a review).

Introduction

A relevant question from a developmental point of view concerns whether these possible audiovisual correspondences have to be learned from experience or else they are already present from very early on, even before language is acquired. Some studies suggest that crossmodal correspondences are rooted in language (see Marks, 1984; Marks, Hammeal, & Bornstein, 1987; Smith & Sera, 1992). These associative links would, according to this hypothesis, emerge after the acquisition of language (that is, after specific space-related lexical items are acquired). In contrast, other studies indicate that language may not be essential for the arising of such crossmodal links. For example, Parkinson, Kohler, Sievers, and Wheatley (2012) explored pitch classification in adults from a hill tribe in Cambodia, whose language does not use spatial-related labels to refer to pitch (as it happens in English, and the vast majority of languages). The perceptual association between pitch and vertical position was present in members of this tribe, suggesting that the presence of space-based terms to describe pitch may not be necessary for this perceptual association to appear (see also Martino & Marks, 1999).

Regarding the development of these crossmodal associations, Walker and colleagues (2010) found evidence of a cross-modal link between dynamic sounds and visuospatial height (as well as between pitch and visual sharpness) in 3- to 4-month-old infants. Infants looked longer at the visual stimuli in the crossmodally congruent conditions (e.g., an ascending frequency sound combined with a visual stimulus moving toward the upper part of the screen) than in the incongruent conditions. Although other authors have failed to replicate these effects (see Lewkowicz & Minar, 2014), Dolscheid, Hunnius, Casasanto,

and Majid (2014) have recently found converging evidence of a crossmodal association between ascending or descending sweeps (low-to-high or high-to-low in frequency) and spatial elevation in 3- to 4-month-old infants, and also between pitch and width. Infants preferred to look at the congruent condition (e.g., a narrow visual stimulus combined with an ascending frequency sweep) than at the incongruent one. In another recent study, Peña, Mehler, and Nespors (2011) also demonstrated that 4-month-old infants are able to perceptually associate ‘high-frontal’ (/i/and/e/) and ‘low-posterior’ (/a/and/o/) vowels with small and large geometric shapes, respectively.

Other perceptual associations between visual and acoustic features, such as size or brightness and pitch, have also been observed in 30–36 month-old children (Mondloch & Maurer, 2004). Haryu and Kajikawa (2012) recently reported that 10-month-old infants associate high/low frequencies with bright/dark objects, respectively. However, no signs of a perceptual relation between pitch and visual size were observed at this age.

In the current study, we further explored the association between pitch and the size of visual stimuli in young infants. In contrast with Peña et al.’s study (2011), we used auditory stimuli (pure-tone frequency sweeps) that bore no relation to human language. This allowed us to address possible perceptual links between visual size and pitch beyond possible “speech-based” (and perhaps more direct) associations (e.g., between different vowels and the size of the mouth aperture). Moreover, in contrast with Haryu and Kajikawa’s (2012) and Mondloch & Maurer’s (2004) studies, dynamic sounds (frequency sweeps) were used instead of low or high frequency flat tones, thus covering a larger spectrum of sound frequencies that ranged between 300 Hz and 1700 Hz (see Dolscheid et al., 2014; Walker et al., 2010). Four- and 6-month-old infants were tested to examine whether this correspondence is already present at these early ages and, therefore, before the acquisition of spatial/abstract lexical items or not (see Swingley, 2009, for a review).

Eighteen full-term 4-month-old infants (9 females; $M = 129$ days, $SD = 5$ days) and 18 full-term 6-month-old infants participated in the study (7 females; $M = 189$ days, $SD = 7$ days). Sixteen additional infants were tested but not included in the final sample because of crying or fussiness (eight 4-month-old and seven 6-month-old infants) and experimental error (one 4-month-old infant). Participants were recruited at the maternity unit at Hospital Sant Joan de Déu. Parental consent was obtained before running the experiment. During the

experiment, the infant sat in a high chair, watching a series of audiovisual animations on a 17-inch TFT monitor. The total amount of time that an infant looked directly to a specific area of the screen was determined by a Tobii T120 Eye-tracker, which was used to collect and store eye-tracking data in an Intel Core i5 computer. Auditory stimuli were delivered from two loudspeakers located at each side of the computer screen.

The experiment contained 4 blocks, randomly presented, each one including 6 trials (24 trials in total). Each block began with an attention getter – a white cross – at the center of the screen, followed by 6 identical trials of 2500 ms of duration. Each trial consisted on two yellow balls appearing simultaneously at the 2 sides (left and right) of the screen, at 15.2 cm from the center (see Fig. 1.). While one of the balls increased in size (from a diameter of 3 cm to a diameter of 13.5 cm), the other one decreased (from a diameter of 13.5 cm to a diameter of 3 cm), at a constant speed of 4.2 cm per second. The presentation of these dynamic visual stimuli was accompanied by an auditory presentation of an ascending (300–1700 Hz) or a descending frequency sweep (1700–300 Hz) at 60 dB (see Fig. 1). The side on which the increasing or decreasing ball appeared changed in a block-by-block basis.

Two areas of interest (AOI) were established at each side of the screen. We calculated the proportion-of-total-looking-time (PTLT) directed at each AOI by dividing the total amount of looking time infants spent at each AOI by the time they spent looking at both AOIs. Additionally, the total amount of looking time spent at each AOI was used for other analyses.

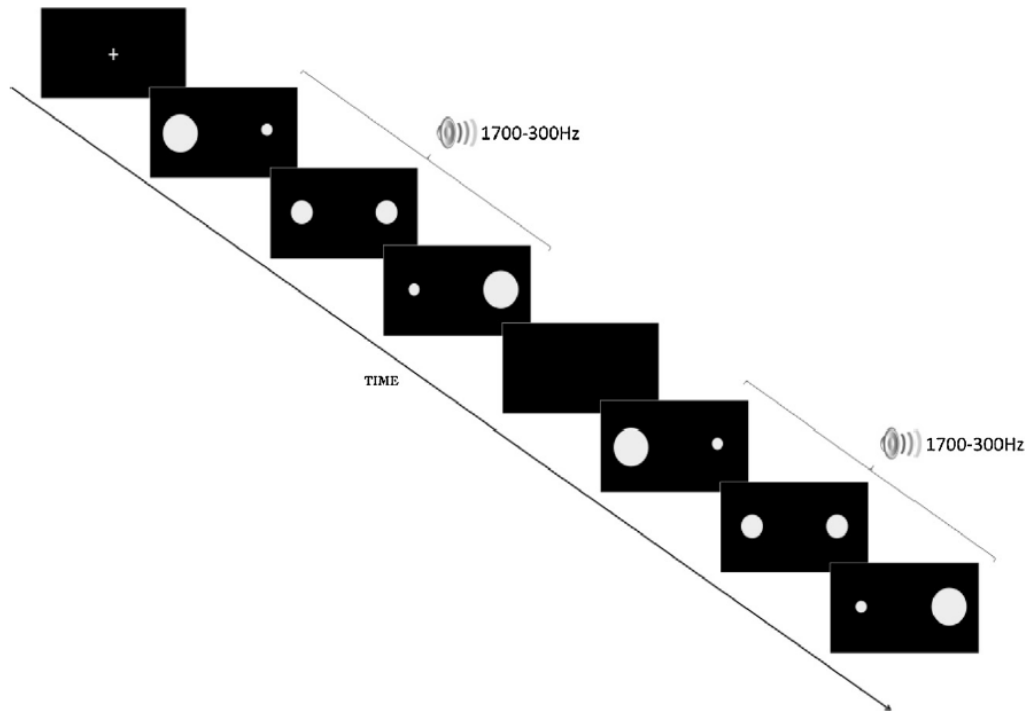


Fig. 1. Example of a sequence of events of two trial animations used in the experiment. The size of the left and the right ball, which were presented in combination with a descending frequency sweep, decreased and increased, respectively.

Analyses including PTLT scores revealed that 6-month-old infants looked at the audiovisually congruent stimuli above chance level (see Fig. 3) [$t(17) = 2.57, p = .02$]. This effect was not observed in the 4-month-old infant group [$t(17) = .37, p = .72$] (see Fig. 2).

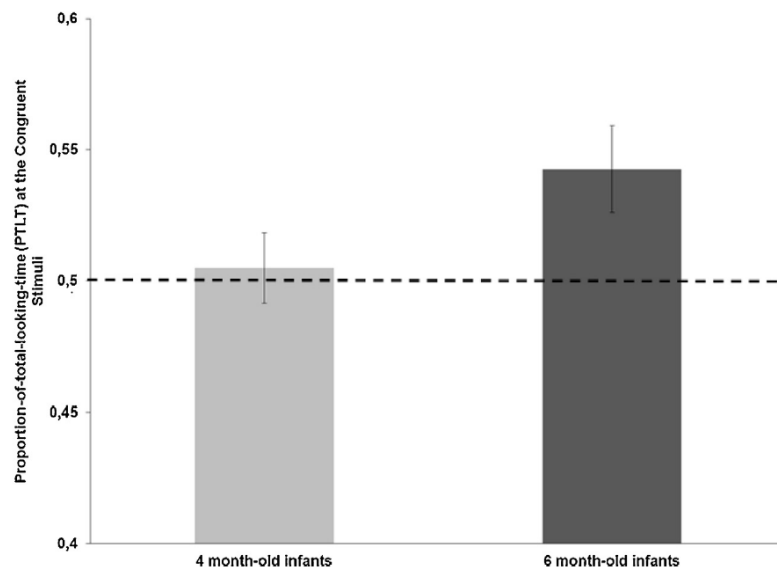


Fig. 2. Proportion-of-total-looking-time (PTLT) for the congruent visual stimuli in 4- and 6-month-old infants. The dotted line indicates chance level. Error bars indicate standard error of the mean.

In order to avoid the dependency between the total looking time to the right and the left AOI within the same trial, we compared the total time looking at both AOI in each trial. That is, we calculated the total amount of time during which infants looked at right and left AOI (right or left ball, respectively) in each trial (note that the AOI's were congruent with the sound in half of the trials, and incongruent in the other half). We then conducted an analysis of variance (ANOVA) including “age” (4-vs. 6-month-old infants) as a between-subject factor, and “congruency” (congruent trials vs. incongruent trials) and “side” (Left vs. Right AOI) as within-subject factors. We found a marginal congruency effect [$F(1, 34) = 3.69, p = .06, 2p = .10$], and a significant interaction between congruency and age [$F(1, 34) = 6.45, p = .02, 2p = .16$]. No interaction between congruency and side was observed [$F(1, 34) = 0.08, p = .78, 2p = .00$].

A follow-up analysis motivated by the interaction found between congruency and age revealed that 6-month-old infants looked longer at the audiovisually congruent trials than at the incongruent ones [$F(17) = 6.44, p = .02, 2p = .27$]. In contrast, 4-month-old infants did not show this looking time difference [$F(17) = .42, p = .52, 2p = .02$] (See Fig. 3).

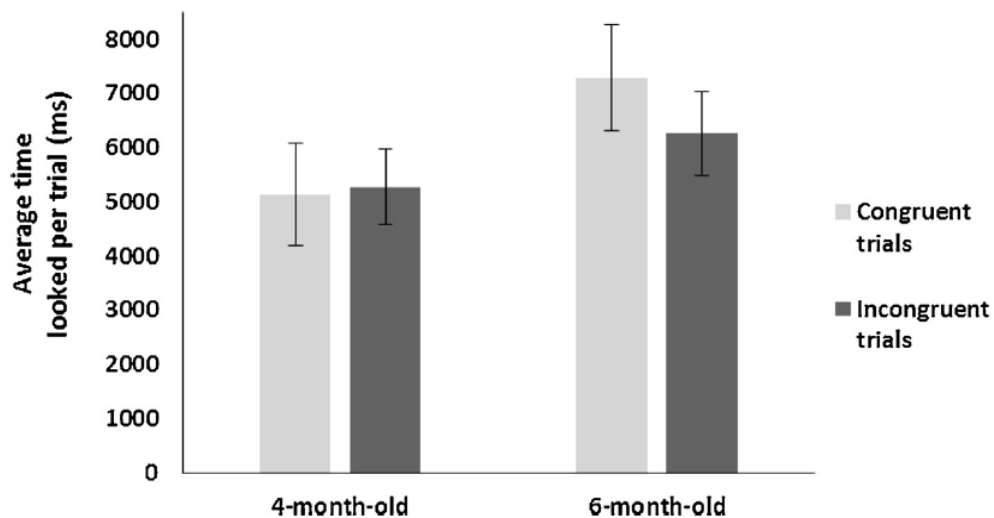


Fig. 3. Average of total looking time in the congruent and incongruent trials at the AOI's in 4- and 6-month-old infants. Bars indicate standard error of the mean.

Taken together, the results of the present study suggest that crossmodal correspondences between visual size and pitch are already present at an early age, being more prominently observed in infants of 6 months of age than in younger (4-month-old) infants. Our findings indicate that the crossmodal association between size and pitch can be observed very early on, but also that some experience (and/or perhaps more cerebral maturation) may be needed for us to observe its presence in infants. Our results reflect that 6 months of exposure to the environment would be enough to consolidate the associative correspondence between visual size and auditory pitch.

Mondloch and Maurer (2004) reported crossmodal links between size (small/big) and pitch in 35-month-old toddlers (see also Marks et al., 1987). Although Mondloch and Maurer's findings are in line with our results, the participants in our study were much younger (and clearly prelexical) infants. Therefore, our results give support to the idea that language is not necessary for crossmodal correspondences to occur. However, this hypothesis does not refute a possible mediation of language in crossmodal correspondences in later stages of life (e.g., by evoking words with spatial or size connotations while performing auditory tasks). The present results are in line with evidence of other crossmodal associations reported both in young infants (e.g., Walker et al., 2010) and in adults whose language does not contain any spatial term to describe pitch (Parkinson et al., 2012). Our results also fit well with previous evidence of a crossmodal association between specific speech sounds and the size of visual objects (Pena et al., 2011). Considering that in the present study frequency-modulated pure tones were used instead of vowels, we can conclude that this crossmodal link is also present in non-speech stimuli, and beyond a plausible articulatory-mediated projection of vowels onto their corresponding size of mouth aperture (see Pena et al., 2011). The present findings may suggest that the crossmodal correspondences between different visuospatial dimensions and pitch do not depend entirely on language.

In contrast with our results, Haryu and Kajikawa (2012) reported that 10-month-old infants were not able to associate pitch and visual size. The apparent discrepancy between the two studies could perhaps be based on methodological issues. In Haryu and Kajikawa's study, static auditory stimuli (a 2096 Hz high-pitched tone and a 262 Hz low-pitched tone) and moving visual stimuli (bouncing balls) were used. Therefore, the intrinsic properties of the auditory and the visual stimuli were different, with the latter being more dynamic than the

former. In our study, both stimuli were dynamic, perhaps increasing the possibility of establishing a link between the auditory and the visual stimuli. The visual and the auditory stimuli could gradually increase or decrease both in size and in frequency. While static tones have been employed in the majority of studies addressing crossmodal connexions (see Chiou & Rich, 2012; Occelli, Spence, & Zampini, 2009; Rusconi et al., 2006), only a few of them have employed frequency sweeps (see Mossbridge, Grabowecky, & Suzuki, 2011; Walker et al., 2010).

Another relevant aspect of our study that might have influenced our results is brightness. As the visual stimulus was increasing in size, the bright surface was also augmenting. Although some studies have shown crossmodal correspondences between brightness and pitch (Haryu & Kajikawa, 2012; Marks, 1974), there is no direct evidence suggesting that the size of brightness' surfaces (with no modification of the level of brightness by itself, as in the present study) can be associated with different sound frequencies. Further research manipulating both brightness and visual size will be needed to investigate how these 2 visual features interact in (perhaps more ecologically-valid) perceptual correspondences.

We believe that our results support the idea that the crossmodal correspondence between pitch and visual size start to arise during the first months of life and is fully acquired at around 6 months of age, after some experience with the environment. At the same time, the crossmodal correspondence between size and pitch seems to be, at these early ages, independent from language.

Acknowledgments

We would like to thank Maria Teixidó for her collaboration in data collection, Sara Rodríguez-Cuadrado for her helpful comments on an earlier version of the article and the infants' families for their participation. The present study was supported by a predoctoral fellowship from Institut de Recerca en Cervell, Cognition & Behaviour (IR3C; Universitat de Barcelona) & Fundació Sant Joan de Déu to I.F.-P. and by grants PSI2012-39149 and PSI2010-20294 from Ministerio de Economía y Competitividad (Spanish Government) to J.N. and F.P., respectively. A fuller report will be provided upon request.

References

- Bien, N., ten Oever, S., Goebel, R., & Sack, A. T. (2012). The sound of size: Crossmodal binding in pitch-size synesthesia: A combined TMS, EEG and psychophysics study. *Neuroimage*: 59., (1), 663–672.
<http://dx.doi.org/10.1016/j.neuroimage.2011.06.095>
- Chiou, R., & Rich, A. N. (2012). Cross-modality correspondence between pitch and spatial location modulates attentional orienting. *Perception*: 41., (3), 339–353.
<http://dx.doi.org/10.1068/p7161>
- Dolscheid, S., Hunnius, S., Casasanto, D., & Majid, A. (2014). Prelinguistic infants are sensitive to space–pitch associations found across cultures. *Psychological Science*: 25., (6), 1256–1261. <http://dx.doi.org/10.1177/0956797614528521>
- Gallace, A., & Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of visual size. *Attention, Perception, & Psychophysics*: 68., (7), 1191–1203. <http://dx.doi.org/10.3758/bf03193720>
- Haryu, E., & Kajikawa, S. (2012). Are higher-frequency sounds brighter in color and smaller in size? Auditory–visual correspondences in 10-month-old infants. *Infant Behavior & Development*: 35., (4), 727–732.
<http://dx.doi.org/10.1016/j.infbeh.2012.07.015>
- Lewkowicz, D. J., & Minar, N. J. (2014). Infants are not sensitive to synesthetic cross-modality correspondences: A comment on Walker et al. (2010). *Psychological Science*, <http://dx.doi.org/10.1177/0956797613516011>
- Marks, L. E. (1974). On associations of light and sound: The mediation of brightness, pitch, and loudness. *American Journal of Psychology*, 173–188. Marks, L. E. (1984). Synesthesia and the arts. *Advances in Psychology*: 19., 427–447.
[http://dx.doi.org/10.1016/S0166-4115\(08\)62362-6](http://dx.doi.org/10.1016/S0166-4115(08)62362-6)
- Marks, L. E., Hammeal, R. J., & Bornstein, M. H. (1987). Perceiving similarity and comprehending metaphor. *Monographs of the Society for Research in Child Development*: 52., (1), 1–102.
- Martino, G., & Marks, L. E. (1999). Perceptual and linguistic interactions in speeded classification: Tests of the semantic coding hypothesis. *Perception*: 28., (7), 903–923. <http://dx.doi.org/10.1068/p2866>

- Melara, R. D., & O'Brien, T. P. (1987). Interaction between synesthetically corresponding dimensions. *Journal of Experimental Psychology: General*: 116., (4), 323–336. <http://dx.doi.org/10.1037/0096-3445.116.4.323>
- Mondloch, C., & Maurer, D. (2004). Do small white balls squeak? Pitch-object correspondences in young children. *Cognitive, Affective, & Behavioral Neuroscience*: 4., (2), 133–136. <http://dx.doi.org/10.3758/CABN.4.2.133>
- Mossbridge, J. A., Grabowecky, M., & Suzuki, S. (2011). Changes in auditory frequency guide visual–spatial attention. *Cognition*: 121., (1), 133–139. <http://dx.doi.org/10.1016/j.cognition.2011.06.003>
- Occelli, V., Spence, C., & Zampini, M. (2009). Compatibility effects between sound frequency and tactile elevation. *Neuroreport*: 20., (8), 793–797. <http://dx.doi.org/10.1097/WNR.1090b1013e32832b38069>
- Parkinson, C., Kohler, P. J., Sievers, B., & Wheatley, T. (2012). Associations between auditory pitch and visual elevation do not depend on language: Evidence from a remote population. *Perception – London*: 41., (7), 854. <http://dx.doi.org/10.1068/p7225>
- Peña, M., Mehler, J., & Nespors, M. (2011). The role of audiovisual processing in early conceptual development. *Psychological Science*: 22., (11), 1419–1421. <http://dx.doi.org/10.1177/0956797611421791>
- Rusconi, E., Kwan, B., Giordano, B. L., Umiltà, C., & Butterworth, B. (2006). Spatial representation of pitch height: The SMARC effect. *Cognition*: 99., 113–129. <http://dx.doi.org/10.1016/j.cognition.2005.01.004>
- Smith, L. B., & Sera, M. D. (1992). A developmental analysis of the polar structure of dimensions. *Cognitive Psychology*: 24., (1), 99–142. [http://dx.doi.org/10.1016/0010-0285\(92\)90004-L](http://dx.doi.org/10.1016/0010-0285(92)90004-L)
- Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*: 73., (4), 971–995. <http://dx.doi.org/10.3758/s13414-010-0073-7>
- Swingle, D. (2009). Contributions of infant word learning to language development. *Philosophical Transactions of the Royal Society B: Biological Sciences*: 364., (1536), 3617–3632. <http://dx.doi.org/10.1098/rstb.2009.0107>
- Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A., & Johnson, S. P. (2010). Preverbal infants' sensitivity to synaesthetic cross-modality

correspondences. *Psychological Science*: 21., (1), 21–25.

<http://dx.doi.org/10.1177/0956797609354734>

5. STUDY 3:

Fernández-Prieto, I., Caprile, C., Tinoco-González, D., Ristol-Orriols, B., López-Sala, A., Póo-Argüelles, P., Pons, F. & Navarra, J. (2016) Pitch perception deficits in nonverbal learning disability. *Research in Developmental Disabilities*, 59, 378-386.

<http://dx.doi.org/10.1016/j.ridd.2016.09.011>

Abstract

The nonverbal learning disability (NLD) is a neurological dysfunction that affects cognitive functions predominantly related to the right hemisphere such as spatial and abstract reasoning. Previous evidence in healthy adults suggests that acoustic pitch (i.e., the relative difference in frequency between sounds) is, under certain conditions, encoded in specific areas of the right hemisphere that also encode the spatial elevation of external objects (e.g., high vs. low position). Taking this evidence into account, we explored the perception of pitch in preadolescents and adolescents with NLD and in a group of healthy participants matched by age, gender, musical knowledge and handedness. Participants performed four speeded tests: a stimulus detection test and three perceptual categorization tests based on colour, spatial position and pitch.

Results revealed that both groups were equally fast at detecting visual targets and categorizing visual stimuli according to their colour. In contrast, the NLD group showed slower responses than the control group when categorizing space (direction of a visual object) and pitch (direction of a change in sound frequency). This pattern of results suggests the presence of a subtle deficit at judging pitch in NLD along with the traditionally-described difficulties in spatial processing.

What does this paper add?

Nonverbal learning disability (NLD) is a complex and multisymptomatic disorder characterized by the presence of deficits in visuospatial processing and in other cognitive abilities such as numerical processing, prosody. However, despite the wide range of cognitive processes affected, auditory perception has been historically discarded as part of the diagnosis criteria for NLD. In the current study, we report a possible deficit in pitch perception in NLD. Our results indicate that children with NLD show significant difficulties at judging the direction of a change in the frequency of sounds. This lower level of pitch perception performance could be associated with structural anomalies previously observed in children with NLD (e.g. white matter alterations in the right hemisphere). The right hemisphere has generally been related to the processing of speech prosody. Keeping in mind that the key mechanism to decode speech prosody is based on deciphering the variations in pitch that are available in the speech signal, our results may perhaps explain the difficulties in understanding speech prosody previously found in NLD.

1. Introduction

Nonverbal learning disability (NLD) is a neurological learning disorder mostly related to the right hemisphere (see Semrud-Clikeman & Hynd, 1990) whose prevalence is estimated to be around 5–10% of all learning disabilities (Davis & Broitman, 2011). This disability was first described by Myklebust (1975) as a subtype of learning disorders that affect non-linguistic abilities.

Although researchers in the field do not question the need of a diagnostic label for NLD, there is some disagreement regarding the criteria used in its differential diagnosis. In an attempt to reduce this ambiguity in the diagnosis of NLD, Mammarella and Cornoldi (2014) (see also Fine, Semrud-Clikeman, Bledsoe, & Musielak, 2013) proposed several criteria based on a review of the last 30 years of research in the field NLD is characterized by (1) poor visuospatial and good verbal intelligence, (2) the presence of difficulties in visuoconstructive and fine-motor abilities, (3) poor mathematical and good reading decoding achievement at school, (4) spatial working memory deficits, and (5) emotional and social difficulties. The authors of this comprehensive study also suggested that the first criterion should always be present in the diagnosis of NLD and at least two of the other four criteria should be met.

So far, the visuospatial and mathematical abilities have monopolized most of the research conducted on NLD. In a recent study, Crollen, Vanderclausen, Allaire, Pollaris, and Noël (2015) observed visuospatial and numerical processing deficits in children diagnosed with NLD. More specifically, these authors explored the spatial representation of numbers (see Dehaene, Bossini, & Giroux, 1993) and found that children with NLD were less able to represent numbers spatially than a control group of healthy children. In particular, children with NLD did not show any tendency to associate small numbers to the left side and bigger numbers to the right side of the external space, a phenomenon previously observed in healthy participants and known as the SNARC effect (Spatial Numerical Association of Response Codes; see Dehaene, Bossini, & Giroux, 1993; see also Hubbard, Piazza, Pinel, & Dehaene, 2005, for a review).

Children and adolescents with NLD do not usually show anomalies in basic language skills (e.g. morphology or phonology), reading decoding, or in any other cognitive function such

as attention or long-term memory (see Mammarella et al., 2009; Pennington, 2009; Rigau-Ratera, Garcia-Nonell, & Artigas-Pallares, 2004; Rourke & Tsatsanis, 2000). Importantly, previous studies have also discarded the presence of auditory deficits in NLD (see Rourke, 1989, 1995).

Regarding the possible neural bases of NLD, it has been suggested the presence of significant white matter perturbations in the right hemisphere of patients with NLD (Rourke, 1987, 1988, 1995). According to Rourke (1987), white matter alterations in the right hemisphere correlate positively with the presence of symptoms commonly described in NLD such as difficulties in visuospatial processing and speech prosody. Given the prominent participation of the right hemisphere (see Gandour et al., 2004; Ross & Monnot, 2008; Tong et al., 2005; Wong, 2002; for a review) and, specially its posterior (parietal) regions (see Perrone-Bertolotti et al., 2013) in processing prosody of speech, it is not surprising that this linguistic dimension is affected in NLD. Shapiro and Danly (1985) demonstrated, after analysing sentences read by patients with lesions in different areas of the brain, that only patients with post-Rolandic posterior lesions in the right hemisphere produce altered prosodic speech characterized with exaggerated variations in pitch. More recently, in a study conducted with sleeping 3-month-old infants, Homae, Watanabe, Nakano, Asakawa, and Taga (2006) showed that the right temporoparietal regions of the brain are, that the right temporoparietal regions of the brain are more sensitive to normal speech – which includes variations in pitch and loudness- than its left counterparts (see also Arimitsu et al., 2011).

Noteworthy, the key mechanism to decode speech prosody is based on decoding variations in pitch; that is, on discriminating dynamic changes in the frequency of sound (see D. Patel, Peretz, Tramo, & Labreque, 1998). Interestingly, patients with a unilateral cerebrovascular lesion in the right hemisphere show poorer perception of pitch contour than patients with the same lesion in the left hemisphere (see Peretz, 1990). Despite the involvement of pitch perception in the comprehension of speech prosody, and the fact that the right hemisphere contributes to both processes, there are no studies specifically exploring possible deficits in pitch perception in patients with NLD.

Another reason for investigating pitch perception in NLD is based on previous behavioural evidence suggesting that, in a similar way as in the already mentioned SNARC effect, high

and low sound frequencies seem to be mapped onto high and low spatial positions, respectively. In fact, there is strong evidence suggesting that the processing of pitch and vertical coordinates in space influence each other in such a way that the responses to one of these dimensions can be modified by manipulating the other (see Melara & O'Brien, 1987; Occelli, Spence, & Zampini, 2009; Rusconi, Kwan, Giordano, Umiltà, & Butterworth, 2006). Crucially for the present study, Melara and O'Brien (1987) demonstrated the existence of a crossmodal link between pitch and spatial elevation. In their study, participants had to make an auditory or visual classification task while ignoring any stimulus variation in the other (irrelevant) dimension. When pitch was the relevant dimension, the participants had to press one of two different buttons every time they perceived a high or a low frequency sound. The auditory stimuli were presented together with a dot that could appear in a higher or lower position with respect of the central point of fixation. When the spatial position was the relevant dimension, the same procedure was used but the participants had to respond to the position of the dot instead of responding to sound pitch. These authors observed that a variation in the auditory dimension affected the participants' responses to the visuospatial dimension and vice versa. Moreover, the results revealed faster responses in the compatible trials, in both auditory and visual conditions (e.g. the dot appeared in a high position while a high frequency sound was presented).

The presence of a strong crossmodal association between the pitch and verticality has also been observed in many studies using a large variety of experimental methodologies (see Occelli et al., 2009; Parise & Spence, 2009; Sonnadara et al., 2009; see Deroy & Spence, 2013; Deroy, Fernández-Prieto, Navarra & Spence, in press; Spence, 2011 for reviews), including indirect tasks in which the participants were not asked to perform any judgment regarding pitch or spatial elevation (Lidji, Kolinsky, Lochy, & Morais, 2007; Rusconi et al., 2006). Perceiving sounds with high or low frequencies can even bias the perceivers' visuospatial attention towards upper or lower spatial locations, respectively (see Chiou & Rich, 2012; Mossbridge, Grabowecky, & Suzuki, 2011).

In the current study, we explored auditory pitch and visuospatial perception in a group of adolescents with NLD (NLD group) and in a control group of adolescents without any neurological or psychiatric disorder. To achieve this goal, four computer-based tests were designed:

1. Auditory pitch categorization test. This test was designed to obtain accurate measures (reaction times, RTs; and accuracy) of the participants' ability to categorize sounds containing either an ascending or a descending change in pitch.
2. Visuospatial categorization test. This test was created as an analogy of the auditory pitch categorization test. The participants' task consisted on judging the direction, along the vertical axis, of a visual stimulus (a filled circle).
3. Speeded stimulus detection test. This test was included in our study to obtain an index of the participants' speed at responding to external stimulation. This was done to ensure that the NLD group was not generally slower in responding to external stimulation than the control group.
4. Colour categorization test. This test was included to address the participants' ability to use sensory information and categorize stimuli according to a specific perceptual dimension (in this case, colour).

While no difference was expected between the two groups in the last two tests (speeded detection and colour categorization), we expected to observe poorer performance in the NLD group than in the control group in the first two tests (involving pitch and visuospatial judgments, respectively). Finally, according to our specific hypothesis that pitch processing would be impaired in NLD, a correlation between the measures obtained in the first two tests was also expected.

2. Method

2.1. Participants

Eight male adolescents with NLD (six right handed, mean age $13.3 \pm SD 1.5$ years) and eight healthy male adolescents (six right handed, mean age $13.4 \pm SD 1.6$ years) participated in the study. The patients with NLD were matched by age and handedness, and were recruited at the Learning Disorders Unit (Unitat de Trastorns de l'Aprenentatge

Escolar, UTAE) of Hospital Sant Joan de Déu (Barcelona, Spain). The patients had already been diagnosed with NLD and were selected according to the following criteria (see Fine, Semrud-Clikeman, Bledsoe, & Musielak, 2013):

- Presence of difficulties in motor abilities (fulfilling the DSM-IV-TR criteria for developmental coordination disorder) as measured in Purdue Pegboard Test (manual dexterity and bimanual coordination). Scores had to be lower than typically developing children by 15 points ($1 < SD$).
- Visuospatial and visuomotor difficulties, with scores lower than typically developing children by 15 points ($1 < SD$) in the NEPSY Arrows subtest and the Blocks Design subtest from the Wechsler Intelligence Scale for Children IV (WISC-IV).
- The presence of at least 2 of the following symptoms:
 - Social difficulties (assessed using the Vineland Adaptive Behavior Scale, VABS). Scores had to be 15 points below typically developing children ($1 < SD$) in communication, living skills and socialization subtests.
 - Verbal IQ higher by 15 points or more than typically developing children ($1 < SD$) in the similarity, vocabulary and comprehension subtests of WISC-IV.
 - Mathematical disability as observed in scores lower than typically developing children by 15 points ($1 < SD$) in the mathematical subtests from WISC-IV.
- Exclusion criteria were the presence of intellectual disability in WISC-IV, the presence of specific learning disorder with reading or written expression impairments (according to Diagnostic and statistical manual of mental disorders, 5th ed., DSM-5), the presence of academic difficulties in linguistic skills, attention deficit/hyperactivity disorder (as assessed by the Conners scales for parents, CPRS-48, and teachers, CTRS-28) and/or autism spectrum disorder (as evaluated by the Autism Diagnostic Interview-Revised, ADI-R).

2.2. Apparatus

An Intel Core computer and a 15-inch CRT monitor (Philips 107-E Monitor, 85 Hz) were used to run the four tests. The experimental procedure was controlled by E-Prime 2.0 (Psychology Software Tools Inc., Pittsburg, PA). The study was conducted in a dimly-lit and sound-attenuated room. The participants sat at an approximate distance of 60 cm from the monitor. Two loudspeakers (Phillips A 1.2 Fun Power, China) were located at each side of the computer screen.

2.3. Experimental tasks

All of the participants performed the four computerized tests. Their performance was measured in terms of reaction times (RTs) in correct responses and accuracy (the proportion of incorrect responses in each test). These were considered as the dependant variables in all of the statistical analyses.

2.3.1. Auditory pitch categorization test

This test was divided in two blocks, with a one-min pause between them. Each block included 44 test trials. The participants had to identify the direction (ascending or descending) of a progressive change in the sound frequency of pure tones. There were four different frequency sweeps with a duration of 2100 ms; two ascending (with a frequency ranging from 466 Hz to 587 Hz, and from 932 Hz to 1174 Hz) and two descending (from 466 Hz to 370 Hz, and 932 Hz to 739 Hz). Before the test, the participants were told that the two directions in frequency (ascending or descending) were produced by two different cartoon characters (e.g., the green cartoon character produced ascending sweeps and the red character produced descending sweeps; see Fig. 1A) and that their task consisted on deciding which of the two characters produced the tone in each trial. This association between the cartoons and the tones was counterbalanced across participants. During the main pitch categorization test, participants responded with two different keys of the computer keyboard to indicate the character that produced the tone just presented (e.g., “a” for the green cartoon 1 and “l” for the red cartoon 2). Visual feedback (indicating a

correct or an incorrect response or a response omission) was provided, during 1500 ms, after each response, or after 2400 ms in case of no response. The participants were also instructed to avoid responding when the tone did not change, which occurred in ten percent of the trials. These ‘catch’ trials were introduced to ensure that participants paid attention to the sounds.

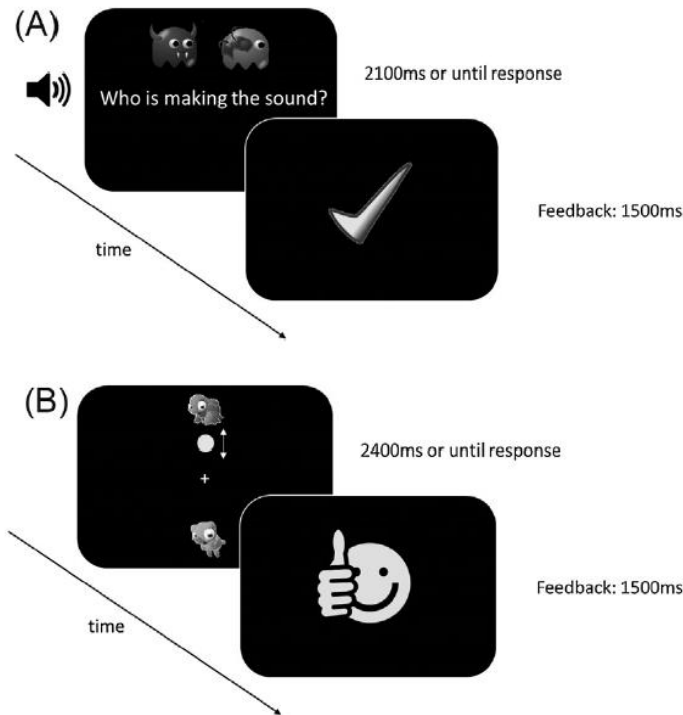


Fig. 1. (A) Example of an experimental trial in the Auditory Pitch Categorization Test. The sound played could be four different frequency sweeps; two ascending and two descending (10% of trials the sound did not change). Participants could receive three different feedbacks after the response: positive (correct response), negative (incorrect response) or warning signal (omission or no response). (B) Example of an experimental trial in the Visuospatial Categorization Test. The circle appeared above or below the fixation point, and could move up or down (10% of trials the circle did not move). Participants could receive three different feedbacks after the response: positive (correct response), negative (incorrect response) or warning signal (omission or no response).

2.3.2. Visuospatial categorization test

The test was divided in two blocks, each one including 77 trials, and also included a one-minute pause between the blocks. The participants had to judge whether a circle (1.5 cm of diameter and located 3 cm above or below of the fixation point) was moving towards a

cartoon image of a turtle (located at the top of the screen; 6 cm above a centrally-presented fixation point), or a cartoon image of a pig (located at the bottom of the screen; 6 cm below the fixation point). Each trial started with the appearance of a fixation white cross (2 cm x 2 cm) in the middle of the screen, on a black background. The circle always appeared slightly above or below the fixation point, and moved up or down in the vertical axis until the participant's response or up to 2400 ms in case of no response. As in the previous test, participants were instructed to press the "a" or "l" keys using the index finger of their left and right hand, respectively. Each of these two keys was associated to one of the cartoons and this association was counterbalanced across participants. In order to ensure that the participants were attending to the visual stimuli, they were encouraged to avoid responding in catch trials where the circle did not move, which represented the 10% of the total amount of trials. As in the previous test, visual feedback was provided for 1500 ms after each participant's response or after 2400 ms if no response was registered (see Fig. 1B).

2.3.3. Speeded stimulus detection test

In 60 different trials, participants had to detect a yellow circle (diameter: 3 cm) and press the keyboard's space bar as quickly as possible with their dominant hand. Each trial started with the appearance of a fixation white cross (2 cm x 2 cm) in the middle of the screen, on a black background. The white cross remained on the screen for a period randomly chosen from the following time intervals: 600, 800, 1000, 1200, 1400 or 1600 milliseconds. After this period, a yellow circle (diameter: 3 cm) was presented on the centre of the computer screen until the participant's response or for 3000 ms, in case of no response.

2.3.4. Colour categorization test

The participant had to judge the colour of a circle (diameter: 3 cm) as fast and accurately as possible in 60 different trials. The procedure was identical to the previous test with the following exceptions: the colour of the circle could be either blue or yellow, and the task consisted on pressing the "a" or "l" keys as quickly as possible using the index finger of their left and right hand, respectively, to indicate the colour of the ball (i.e., blue or

yellow). The association between the response keys and the colours was counterbalanced across participants.

3. Results

Considering the small size of the sample of participants, a non-parametric statistical approach was adopted to analyse the data obtained in all of the four tests included in the study. It is worth highlighting that using cognitive computer-based methods allows us to obtain reliable (and generalizable) data even in small sample sizes such as the one included in the current study. This is so because of (1) the capability of computerized tasks to deliver highly-accurate measures (e.g., at the level of milliseconds) of cognitive functions, and (2) the fact that the measures that are obtained in these kind of tasks are based on sensory-driven speeded judgments, rather than based on other perhaps more subjective decisions. Another key factor that increases statistical power, both in the present and in many previous studies, is the fact that the participants' measures are also obtained from an average of many different data points (e.g., 154 trials, in one particular test).

3.1. Auditory pitch categorization test

The NLD and the control group were compared in terms of RTs from trials with a correct response, omission errors (i.e., not responding), categorization errors (i.e., choosing the wrong cartoon character), and false alarms in catch trials (i.e., pressing a key instead of inhibiting response for tones without pitch change). A non-parametric Mann–Whitney U test conducted with RT data revealed slower responses in the NLD group than in the control group ($U = -3.151$, $p = 0.001$; see Fig. 2A). The NLD group also showed a significantly lower percentage of correct responses than the control group ($U = 3.002$, $p = 0.001$). Further analyses on error responses revealed poorer performance in the NLD group compared to the control group in all of the 3 types of error (categorization: $U = 0.268$, $p = 0.021$; false alarms: $U = -2.409$, $p = 0.015$; and omissions: $U = -3.424$, $p < 0.001$) (Table 1).

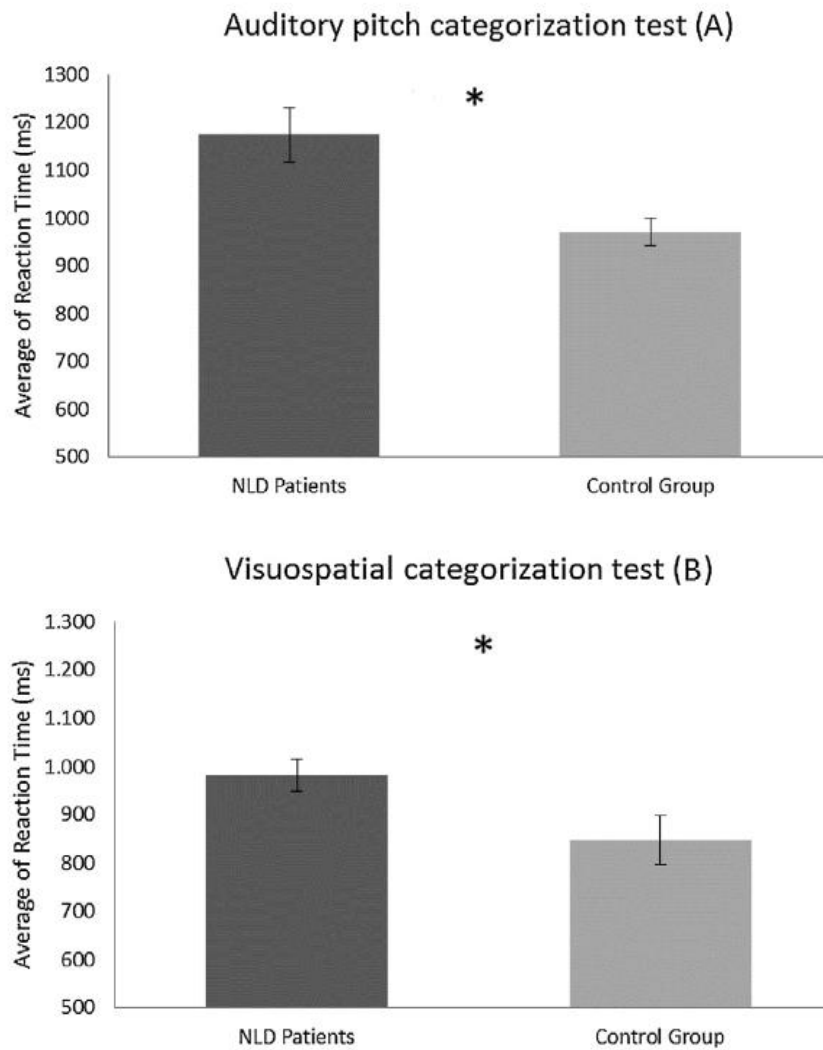


Fig. 2. Average of Reaction Times (in milliseconds) in the two tests used in the study to measure visuospatial and pitch processing. (A) Average of RTs in the Visuospatial Categorization Test in NLD Patients and in the Control group. (B) Average of RTs in the Auditory Pitch Categorization Test in NLD Patients and in the Control group. Error bars show standard errors. Asterisks show significant differences between groups.

Table 1. Percentage of errors and standard deviations of the NLD and the Control group in the four computerized tests.

		Categorization errors	Omission errors	False alarms (catch trials)*
Auditory pitch categorization test	NLD Group	17.5% (SD= 11.28%)	12.34% (SD= 6.56%)	56.28% (SD= 34.72%)
	Control Group	5.78% (SD= 4.32%)	0.78% (SD= 0.65%)	10.94% (SD= 12.39%)
Visuospatial categorization test	NLD Group	3.68% (SD= 3.11%)	0.37% (SD= 0.68%)	3.46% (SD= 5.32%)
	Control Group	5.29% (SD= 4.21%)	0.45% (SD= 0.37%)	10.56% (SD= 7.73%)
Speeded stimulus detection test	NLD Group		8.33% (SD = 2.95%)	
	Control Group		7.70% (SD = 4.27%)	
Colour categorization test	NLD Group	13.95% (SD = 5.19%)	5.63% (SD = 2.66%)	
	Control Group	13.96% (SD = 5.84%)	7.29% (SD = 3.56%)	

* The percentage of false alarms was calculated based on catch trials where participants had to avoid responding (10% of the total amount of total trials).

3.2. Visuospatial categorization test

The NLD and the control group were compared in terms of RTs from trials with a correct response, omission errors, categorization errors (i.e., choosing the wrong direction) and false alarms in catch trials (where participants had to avoid responding). A non-parametric U Mann-Whitney analysis revealed that NLD group's responses at judging the circle movement direction of the circle were significantly slower than the control group ($U = -2.100$, $p = 0.038$; see Fig. 2B). No significant differences were found between the two groups in terms of correct responses percentage ($U = -0.686$, $p = 0.505$), the percentage of categorization or omission errors (categorization: $U = -0.527$, $p = 0.645$; and omissions: $U = -0.817$, $p = 0.051$). The control group showed a significantly higher percentage of false alarms in catch trials than the group of adolescents with NLD ($U = -2.014$, $p = 0.05$) (see Table 1).

3.3. Speeded stimulus detection test

Reaction times, obtained from trials with a correct response, and omission errors (i.e., not responding after the appearance of the visual target) were used to compare the performance in the two groups (NLD and control participants). A non-parametric Mann-Whitney U

test revealed equivalent RTs in both groups ($U = 9.522$, $p = 1$) (see Fig. 3A). No significant differences were observed in terms of omission errors between groups ($U = -0.214$, $p = 0.878$; see Table 1).

3.4. Colour categorization test

The NLD and the control group were compared in terms of RTs in trial with a correct response, omission errors and categorization errors (i.e., choosing the wrong colour). A non-parametric Mann-Whitney U showed no difference in RTs between the two groups ($U = -0.105$, $p = 0.959$) (see Fig. 3B). No difference was found between the two groups in terms of omission or categorization errors ($U = -0.691$, $p = 0.505$; $U = 0.643$, $p = 0.574$; respectively) (see Table 1).

3.5. Correlation between visuospatial and pitch categorization

The non-parametric Spearman's correlation test was used to evaluate the possible relationship between the participants' performance in the principal measures of the present study, mainly the ones obtained in the visuospatial categorization test and the auditory pitch categorization test; including RTs and errors (i.e., omissions, false alarms and categorization errors). Considering data from the whole sample (NLD patients and the Control group), a significant correlation was found between the RTs in these two tests ($R = 0.679$, $p = 0.004$). Participants that were slower at categorizing pitch were also slower at categorizing the direction of a movement. However, no correlation was found between the two tests in any type of error (omission: $R = -0.117$, $p = 0.667$; false alarms: $R = -0.082$, $p = 0.762$ and categorization error: $R = 0.006$, $p = 0.983$).

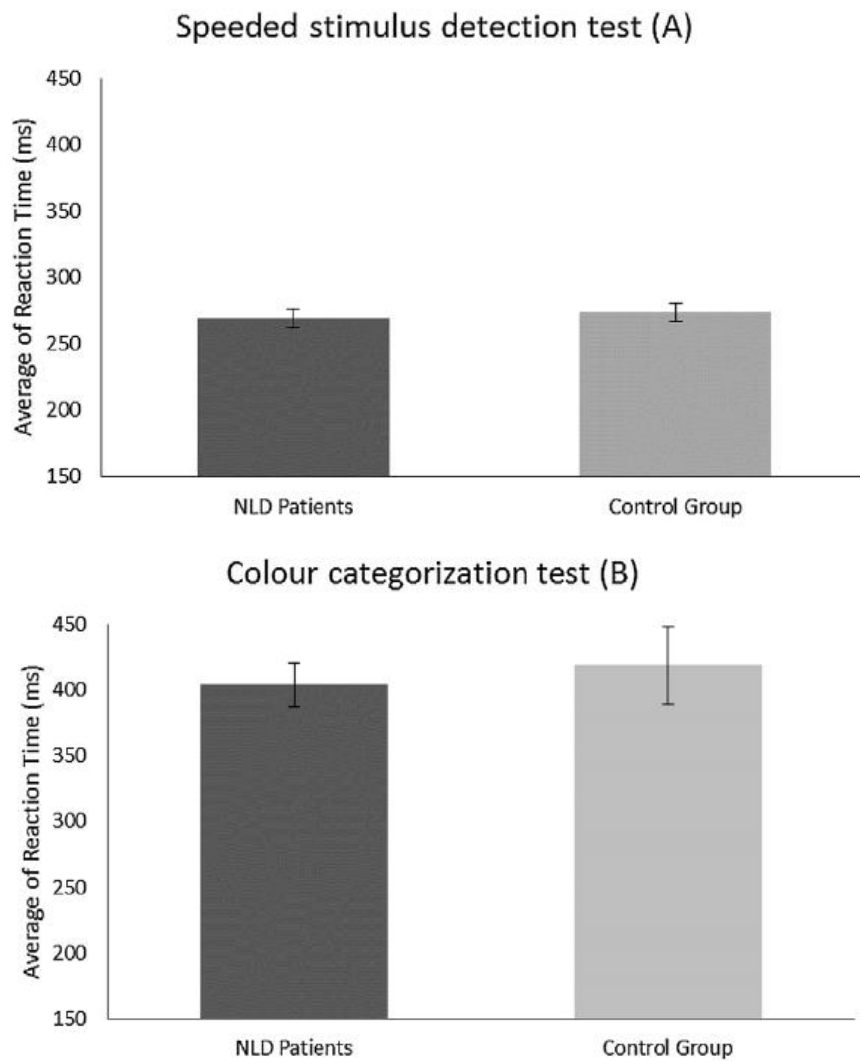


Fig. 3. Average of Reaction Times (in milliseconds) in the two control tests used in the study to assess participants' speed at responding to external stimulation: (A) Average of RTs in the Speeded Detection Test in the group of adolescents with NLD and in the Control group. (B) Average of RTs in the Colour Categorization Test in the NLD Patients and in the Control group. Error bars show standard errors. Asterisks show significant differences between groups.

4. Discussion

Taken together, our results indicate that both the NLD and the control group performed equivalently in a simple stimulus detection test and a colour categorization test. This pattern of results suggests that the participants with NLD did not show general slowness or poorer processing (i.e., leading to an increase of errors) when responding to external stimulation. Importantly, the participants included in this group were able to categorize

stimuli according to a non-spatial dimension (i.e., colour). In contrast, the results of the visuospatial test confirmed our hypotheses: The NLD participants were slower at responding to the direction of a visual object's movement. Crucially, the results obtained in the pitch categorization test indicate that the participants with NLD also had a poorer performance (observed both in RTs and in the percentage of number of errors) at judging the direction of a dynamic change in auditory pitch (ascending vs. descending). Finally, a positive correlation was found, in the participants' RTs, between the results obtained in the visuospatial categorization and the auditory pitch categorization test.

Despite the fact that the computerized tasks provided us with statistically reliable data, a limitation of the present study is the small sample tested. We tried to compensate this restriction with (1) a strict match between the two groups (in terms of age, gender, handedness and musical knowledge), (2) the use of non-parametric analyses, and, perhaps more importantly, (3) the use of highly-accurate computerized measures of cognitive functions based on many data points. Another possible limitation of the study is the difficulty of obtaining a differential diagnosis of NLD. Although the participants in the current study were carefully selected among a larger sample of NLD patients (that included children with comorbidities) and did not meet the criteria for specific learning disorder with impairment in reading and/or with impairment in written expression (see Diagnostic and statistical manual of mental disorders, 5th ed., DSM-5), we cannot rule out, in absence of a closer examination of language skills, that they did not show any linguistic difficulty that remained undetected during their clinical assessment at our centre. Further testing, perhaps including a larger sample of participants with NLD, will be needed to elucidate whether the subtle deficits at judging pitch observed in the present study correlate with specific linguistic abilities involving the adequate processing of variations in sound frequency (e.g., prosody and intonation; see Rourke, 1995).

Children and adolescents with specific language impairment (SLI) often show difficulties in other cognitive areas such as visuomotor abilities (Hill, 2001), social skills (Beitchman et al., 1996) and/or mathematics (Mainela-Arnold, Alibali, Ryan, & Evans, 2011). Previous studies have even associated SLI with an abnormal development of brain structures underlying the procedural memory system (affecting learning, motor and other cognitive skills; see Ullman & Pierpont, 2005). Despite of the relative overlap between SLI and NLD in terms of specific symptoms such as motor and abstract reasoning skills, it is

worth highlighting that the NLD patients that participated in the present study did not show any obvious difficulty in language (as observed both at school and during the psychological and psychiatric assessment conducted at our centre) and had normal verbal intelligence (see Fine, Semrud-Clikeman, Bledsoe, & Musielak, 2013).

To our knowledge, there is no previous evidence suggesting the presence of basic auditory deficits at the level pitch categorization in NDL. In our opinion, although the children with NLD can present short-term memory impairments when they need to bind information from different categories (e.g. shape and colour), these difficulties seem to appear in more complex tasks including a large number of combinations. In a study conducted by Garcia, Mammarella, Pancera, Galera, and Cornoldi (2015), children with NLD showed poorer performance than a control group in memory for shape-colour bindings, the task had a total of 72 combinations (8 shapes and 9 colours). In stark contrast with this previous work, our much easier auditory categorization task had only two possible combinations (e.g. each of two cartoon characters combined with two possible sounds), imposing much less short-term memory demands to the participants.

Non-verbal learning disorder has mainly been characterized, among other aspects, by the presence of difficulties in spatial and abstract tasks (Crollen et al., 2015; Fine, 2012; Rourke, 1995). Our results support and complement this previous evidence by revealing the presence of anomalies in spatial-driven judgments regarding the direction of a visual object's movement. Our results also extend previous evidence reflecting perceptual and cognitive anomalies in NLD to the much less studied case of auditory perception, where difficulties in judging pitch have been found. Due to their difficulties to decode speech prosody (see Rourke, 1989), NLD patients are less able to produce and understand language intonation, thus having difficulties to grasp the difference between statements, commands, questions, etc. (see Rourke, 1995). An adequate processing of pitch variations in speech is essential to understand the others' messages (see Patel et al., 1998). At a speculative level, the difficulties in understanding speech prosody previously reported in the literature may easily be attributed to a deficit at the level of pitch representation. Regarding the possible neural basis of perceptual anomalies reported in the present study, several studies conducted in adults revealed the principal involvement of the right hemisphere in the processing of speech prosody (i.e., processing the pitch and loudness contour in speech signals) (see Gandour et al., 2004; Tong et al., 2005). This is particularly

true for posterior (parietal) regions of this hemisphere (see Perrone-Bertolotti et al., 2013). Therefore, our study, together with previous evidence suggesting the presence of structural anomalies in the right hemisphere in children with NLD (Rourke, 1988) and the involvement of this hemisphere in the processing of speech prosody (see; Shapiro and Danly; 1985; Homae et al., 2006; Arimitsu et al., 2011) may perhaps indicate that deficits found in the perception of prosody may be the consequence of a more basic deficit in perceiving pitch. Previous neuroimaging studies conducted in healthy adults suggest that detection of subtle pitch variations embedded in a melody significantly activates the intraparietal sulcus (IPS) (Foster & Zatorre, 2010a, 2010b), an area considered to be involved in spatial-based tasks. If some specific pitch-based tasks are linked with spatial processing (see Foster & Zatorre, 2010a, 2010b), it seems plausible that a structural and/or functional impairment of these brain areas of the right hemisphere (as it is reported in NLD patients; see Rourke, 1988) may affect both, pitch and visuospatial processing.

According to Rourke (1988) and his “white matter” account of NLD, this pathology could be the result of white matter abnormalities in the right hemisphere. However, further research is needed, perhaps using functional magnetic resonance (fMRI), to relate the auditory and visuospatial deficits found in the present study to possible dysfunctional areas of the brain. The visuospatial and the pitch categorization tasks were, in the current study, designed to avoid the use of language with spatial connotations. Labels such as “arriba” (up), “abajo” (down), “alto” (high), or “bajo” (low) were avoided both in the instructions given to participants and during the test sessions. Participants were instructed to respond to a cartoon associated to the response (e.g., the ascending sound was produced by the monster A and descending sound by the monster B). Furthermore, participants of the current study were Spanish and Catalan speakers, and neither of these two languages use spatial terms in their most frequent words to describe pitch (“agudo”, in Spanish, and “agut”, in Catalan, for high-pitched sounds and “grave”, in Spanish, and “greu”, in Catalan, for low-pitched sounds). So far, the possible role of language labelling in the experimental demonstrations of crossmodal correspondences between space and pitch has been controversial (see Spence, 2011; for a review), especially when taking into account the strong evidence suggesting that these correspondences occur even in prelinguistic infants (Dolscheid, Hunnius, Casasanto, & Majid, 2014; Walker et al., 2010).

In summary, our results show, in line with previous studies, that children with NLD present difficulties in a visuospatial task. This is, according to previous literature, the core of the cognitive deficits that characterize this neurological disorder. In addition, and more importantly, our study also revealed a slower and poorer performance in a pitch-based auditory task in these patients. This difficulty in categorizing an auditory change in frequency could be associated with structural anomalies previously observed in NLD. More research is needed to clarify this possible association. The deficit in pitch processing could perhaps underlie other difficulties in other cognitive areas (e.g., understanding and producing speech prosody in face-to-face interaction) already found in patients with NLD. Therefore, our results could open new paths in the diagnosis and treatment of this cognitive disability.

Acknowledgements

We would like to thank the participants and their families for their participation. The present study was supported by a predoctoral fellowship from Institut de Recerca en Brain, Cognition & Behaviour (IR3C; Universitat de Barcelona) & Fundació Sant Joan de Déu to I.F.-P. and by grants PSI2012-39149 and PSI2010-20294 from Ministerio de Economía y Competitividad (Spanish Government) to J.N. and F.P., respectively.

References

- Arimitsu, T., Uchida-Ota, M., Yagihashi, T., Kojima, S., Watanabe, S., Hokuto, I., et al. (2011). Functional hemispheric specialization in processing phonemic and prosodic auditory changes in neonates. *Frontiers in Psychology*, 2
<http://dx.doi.org/10.3389/fpsyg.2011.00202>
- Beitchman, J. H., Wilson, B., Brownlie, E. B., Walters, H., Inglis, A., & Lancee, W. (1996). Long-term consistency in speech/language profiles: II. Behavioral, emotional, and social outcomes. *Journal of the American Academy of Child & Adolescent Psychiatry*, 35(6), 815–825. <http://dx.doi.org/10.1097/00004583-199606000-00022>
- Crollen, V., Vanderclausen, C., Allaire, F., Pollaris, A., & Noël, M.-P. (2015). Spatial and numerical processing in children with non-verbal learning disabilities. *Research in Developmental Disabilities*, 47, 61–72.
<http://dx.doi.org/10.1016/j.ridd.2015.08.013>
- Davis, J. M., & Broitman, J. (2011). *Nonverbal learning disabilities in children: Bridging the gap between science and practice*. Springer Science & Business Media.
- Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, 122(3), 371–396. <http://dx.doi.org/10.1037/0096-3445.122.3.371371-396>
- Deroy, O., & Spence, C. (2013). Why we are not all synesthetes (not even weakly so). *Psychonomic Bulletin and Review*, 20(4), 643–664.
<http://dx.doi.org/10.3758/s13423-013-0387-2>
- Deroy, O., Fernandez-Prieto, I., Navarra, J., & Spence, C. (in press) Unravelling the paradox of spatial pitch. To appear in T. L. Hubbard (Eds.), *Spatial biases in perception and cognition*. Cambridge, UK: Cambridge University Press.
- Dolscheid, S., Hunnius, S., Casasanto, D., & Majid, A. (2014). Prelinguistic infants are sensitive to space-pitch associations found across cultures. *Psychological Science*, 25(6), 1256–1261. <http://dx.doi.org/10.1177/0956797614528521>
- Fine, J. G., Semrud-Clikeman, M., Bledsoe, J. C., & Musielak, K. A. (2013). A critical review of the literature on NLD as a developmental disorder. *Child*

- Neuropsychology, 19(2), 190–223.
<http://dx.doi.org/10.1080/09297049.2011.648923>
- Foster, N. E., & Zatorre, R. J. (2010a). Cortical structure predicts success in performing musical transformation judgments. *Neuroimage*, 53(1), 26–36.
<http://dx.doi.org/10.1016/j.neuroimage.2010.06.042>
- Foster, N. E., & Zatorre, R. J. (2010b). A role for the intraparietal sulcus in transforming musical pitch information. *Cerebral Cortex*, 20(6), 1350–1359.
<http://dx.doi.org/10.1093/cercor/bhp199>
- Gandour, J., Tong, Y., Wong, D., Talavage, T., Dzemidzic, M., Xu, Y., et al. (2004). Hemispheric roles in the perception of speech prosody. *Neuroimage*, 23(1), 344–357. <http://dx.doi.org/10.1016/j.neuroimage.2004.06.004>
- Garcia, R. B., Mammarella, I. C., Pancera, A., Galera, C., & Cornoldi, C. (2015). Deficits in visual short-term memory binding in children at risk of non-verbal learning disabilities. *Research in Developmental Disabilities*, 45, 365–372.
<http://dx.doi.org/10.1016/j.ridd.2015.07.035>
- Hill, E. L. (2001). Non-specific nature of specific language impairment: A review of the literature with regard to concomitant motor impairments. *International Journal of Language & Communication Disorders*, 36(2), 149–171.
<http://dx.doi.org/10.1080/13682820010019874>
- Homae, F., Watanabe, H., Nakano, T., Asakawa, K., & Taga, G. (2006). The right hemisphere of sleeping infant perceives sentential prosody. *Neuroscience Research*, 54(4), 276–280. <http://dx.doi.org/10.1016/j.neures.2005.12.006>
- Hubbard, E. M., Piazza, M., Pinel, P., & Dehaene, S. (2005). Interactions between number and space in parietal cortex. *Nature Reviews Neuroscience*, 6(6), 435–448.
<http://dx.doi.org/10.1038/nrn1684>
- Mainela-Arnold, E., Alibali, M. W., Ryan, K., & Evans, J. L. (2011). Knowledge of mathematical equivalence in children with specific language impairment: insights from gesture and speech. *Language, Speech, and Hearing Services in Schools*, 42(1), 18–30. [http://dx.doi.org/10.1044/0161-1461\(2010/09-0070\)](http://dx.doi.org/10.1044/0161-1461(2010/09-0070))
- Mammarella, I. C., & Cornoldi, C. (2014). An analysis of the criteria used to diagnose children with Nonverbal Learning Disability (NLD). *Child Neuropsychology*, 20(3), 255–280. <http://dx.doi.org/10.1080/09297049.2013.796920>

- Mammarella, I. C., Meneghetti, C., Pazzaglia, F., Gitti, F., Gomez, C., & Cornoldi, C. (2009). Representation of survey and route spatial descriptions in children with nonverbal (visuospatial) learning disabilities. *Brain and Cognition*, 71(2), 173–179. <http://dx.doi.org/10.1016/j.bandc.2009.05.003>
- Melara, R. D., & O'Brien, T. P. (1987). Interaction between synesthetically corresponding dimensions. *Journal of Experimental Psychology: General*, 116(4), 323. <http://dx.doi.org/10.1037/0096-3445.116.4.323>
- Mossbridge, J. A., Grabowecky, M., & Suzuki, S. (2011). Changes in auditory frequency guide visual-spatial attention. *Cognition*, 121(1), 133–139. <http://dx.doi.org/10.1016/j.cognition.2011.06.003>
- Myklebust, H. R. (1975). . *Progress in learning disabilities* (Vol. 3) New York, NY: Grune & Stratton.
- Occelli, V., Spence, C., & Zampini, M. (2009). Compatibility effects between sound frequency and tactile elevation. *Neuroreport*, 20(8), 793–797, 710.1097/WNR.1090b1013e32832b38069.
- Patel, A. D., Peretz, I., Tramo, M., & Labreque, R. (1998). Processing prosodic and musical patterns: A neuropsychological investigation. *Brain and Language*, 61(1), 123–144. <http://dx.doi.org/10.1006/brln.1997.1862>
- Pennington, B. F. (2009). Nonverbal learning disability. In *Diagnosing learning disorders: A neuropsychological framework* (2nd ed., pp. 242–248). New York, NY: The Guildford Press.
- Peretz, I. (1990). Processing of local and global musical information by unilateral brain-damaged patients. *Brain*, 113(4), 1185–1205. <http://dx.doi.org/10.1093/brain/113.4.1185>
- Perrone-Bertolotti, M., Dohen, M., Loevenbruck, H., Sato, M., Pichat, C., & Baciú, M. (2013). Neural correlates of the perception of contrastive prosodic focus in French: A functional magnetic resonance imaging study. *Human Brain Mapping*, 34(10), 2574–2591. <http://dx.doi.org/10.1002/hbm.22090>
- Rigau-Ratera, E., Garcia-Nonell, C., & Artigas-Pallares, J. (2004). Características del trastorno de aprendizaje no verbal. *Revista de Neurología*, 38(Suppl 1), S33–38.
- Ross, E. D., & Monnot, M. (2008). Neurology of affective prosody and its functional-anatomic organization in right hemisphere. *Brain and Language*, 104(1), 51–74. <http://dx.doi.org/10.1016/j.bandl.2007.04.007>

- Rourke, B. P., & Tsatsanis, K. D. (2000). Nonverbal learning disabilities and Asperger syndrome. *Asperger Syndrome*, 231–253.
- Rourke, B. P. (1987). Syndrome of nonverbal learning disabilities: The final common pathway of white-matter disease/dysfunction? *Clinical Neuropsychologist*, 1(3), 209–234. <http://dx.doi.org/10.1080/13854048708520056>
- Rourke, B. P. (1988). The syndrome of nonverbal learning disabilities: Developmental manifestations in neurological disease, disorder, and dysfunction. *Clinical Neuropsychologist*, 2(4), 293–330. <http://dx.doi.org/10.1080/13854048808403271>
- Rourke, B. P. (1989). *Nonverbal learning disabilities: The syndrome and the model*. New York, NY: Guilford Press.
- Rourke, B. P. (1995). *Syndrome of nonverbal learning disability: Neurodevelopmental manifestations*. New York, NY: Guilford Press.
- Rusconi, E., Kwan, B., Giordano, B. L., Umiltà, C., & Butterworth, B. (2006). Spatial representation of pitch height: The SMARC effect. *Cognition*, 99(2), 113–129. <http://dx.doi.org/10.1016/j.cognition.2005.01.004>
- Semrud-Clikeman, M., & Hynd, G. W. (1990). Right hemisphere dysfunction in nonverbal learning disabilities: Social, academic, and adaptive functioning in adults and children. *Psychological Bulletin*, 107(2), 196–209. <http://dx.doi.org/10.1037/0033-2909.107.2.196>
- Shapiro, B. E., & Danly, M. (1985). The role of the right hemisphere in the control of speech prosody in propositional and affective contexts. *Brain and Language*, 25(1), 19–36. [http://dx.doi.org/10.1016/0093-934X\(85\)90118-X](http://dx.doi.org/10.1016/0093-934X(85)90118-X)
- Sonnadara, R. R., Gonzalez, D. A., Hansen, S., Elliott, D., & Lyons, J. L. (2009). Spatial properties of perceived pitch: influence on reaching movements. *Annals of the New York Academy of Sciences*, 1169, 503–507. <http://dx.doi.org/10.1111/j.1749-6632.2009.04858.x>
- Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, 73(4), 971–995. <http://dx.doi.org/10.3758/s13414-010-0073-7>

- Tong, Y., Gandour, J., Talavage, T., Wong, D., Dziedzic, M., Xu, Y., et al. (2005). Neural circuitry underlying sentence-level linguistic prosody. *Neuroimage*, 28(2), 417–428. <http://dx.doi.org/10.1016/j.neuroimage.2005.06.002>
- Ullman, M. T., & Pierpont, E. I. (2005). Specific language impairment is not specific to language: The procedural deficit hypothesis. *Cortex*, 41(3), 399–433. [http://dx.doi.org/10.1016/S0010-9452\(08\)70276-4](http://dx.doi.org/10.1016/S0010-9452(08)70276-4)
- Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A., et al. (2010). Preverbal infants' sensitivity to synaesthetic cross-modality correspondences. *Psychological Science*, 21(1), 21–25. <http://dx.doi.org/10.1177/0956797609354734>
- Wong, P. C. M. (2002). Hemispheric specialization of linguistic pitch patterns. *Brain Research Bulletin*, 59(2), 83–95. <http://dx.doi.org/10.1016/S0361-923>

6. DISCUSIÓN

A través de los diferentes estudios expuestos en la presente disertación podemos confirmar que existe una asociación transmodal robusta entre determinados atributos perceptivos visuoespaciales (por ejemplo, el tamaño o la elevación en el espacio) y la altura tonal. Dichas asociaciones parecen desarrollarse a lo largo de la vida, aunque, cómo pudimos observar en el estudio 2, ya se manifiestan en edades muy tempranas y parecen consistentes en adolescentes (grupo control, Estudio 3) y adultos (Estudio 1), además de verse afectadas en caso de alteración funcional (y/o estructural) de áreas cerebrales tradicionalmente asociadas al procesamiento de dimensiones visuoespaciales (Estudio 3).

Los resultados obtenidos están en línea con estudios recientes sobre los mecanismos cerebrales implicados en el procesamiento de la altura tonal (Foster y Zatorre, 2010a, 2010b). La confirmación de que la altura tonal es procesada no sólo a nivel auditivo, sino también a nivel espacial, representaría una clara evidencia de la existencia de asociaciones transmodales.

Nuestros datos sugieren que existiría un primer análisis temprano a nivel auditivo de la altura tonal, para posteriormente emerger cierta representación visuoespacial de este estímulo auditivo, es decir, el sonido se situaría en un sistema de coordenadas mental (Walsh, 2003), pudiendo asignar a dicho sonido otras cualidades visuoespaciales, como por ejemplo el tamaño o la angulosidad. Siguiendo trabajos previos tanto teóricos (Walsh, 2003) como experimentales (Foster y cols., 2010a), la presentación de un tono no solo activaría determinadas áreas de la corteza auditiva, sino que también activaría áreas parietales específicas asociadas a localización espacial o al análisis del tamaño (por ejemplo, el surco intraparietal, IPS). Los hallazgos descritos en la presente tesis doctoral sugieren que el procesamiento temprano de estímulos auditivos vendría acompañado de un procesamiento posterior, más "supramodal" si se quiere, que implicaría la activación de representaciones compartidas con el sistema visuoespacial.

A continuación, se procederá a discutir los resultados derivados de cada uno de los tres estudios que forman la presente disertación doctoral.

6.1 Modulación de la atención a través de asociaciones transmodales en adultos

La asociación transmodal entre la altura tonal y la localización espacial es un fenómeno que ha sido demostrado en diversos estudios (Spence, 2011). En cambio, el papel que dicha asociación puede jugar en la modulación de la atención espacial todavía no está claro (Chiou y Rich, 2012; Mossbridge, Grabowecky, y Suzuki, 2011).

Los resultados del Estudio 1 clarifican la influencia de la posible representación espacial asociada a la altura tonal. ¿En qué medida la presentación de un estímulo auditivo específico puede modular la detección de estímulos visuales en el espacio?

En línea con los resultados reportados por Chiou y Rich (2012), observamos un efecto significativo general de señalización espacial transmodal, es decir, las señales auditivas congruentes con los targets visuales (por ejemplo, la presentación de un sonido ascendente y la posterior aparición de un asterisco en uno de los recuadros superiores) producían respuestas de detección más rápidas. Sin embargo, nuestros datos mostraron diferencias significativas entre los efectos producidos por sonidos ascendentes y descendentes en la capacidad de modular la detección de un estímulo visual en el espacio. De acuerdo con nuestra hipótesis, únicamente los sonidos ascendentes produjeron un efecto transmodal, es decir, los participantes fueron significativamente más rápidos detectando el asterisco en los recuadros superiores que en los inferiores cuando su aparición era precedida por un sonido ascendente. En cambio, los sonidos descendentes no produjeron efectos en la velocidad de detección del estímulo visual en los recuadros inferiores comparado con los recuadros superiores.

Este resultado resulta novedoso, ya que la gran mayoría de estudios transmodales entre el sonido y la elevación espacial no informan de posibles diferencias de efecto espacial entre los sonidos graves y los sonidos agudos.

Varios estudios han mostrado que los sonidos agudos generan una mayor respuesta fisiológica comparada con los sonidos graves (Näätänen, 1990; Näätänen, Gaillard, y Mäntysalo, 1978; Ruusuvirta y Astikainen, 2012). Además, parece existir cierta preferencia del sistema perceptivo por los sonidos más agudos. En la infancia, el procesamiento de sonidos agudos en el sistema auditivo se desarrolla previamente al procesamiento de sonidos más graves (Olsho, 1984; Olsho, Koch, y Halpin, 1987; Trehub, Schneider, y Endman, 1980). A nivel atencional, las frecuencias altas incrementan el nivel de alerta (Tomatis, 1978). A nivel psicofisiológico, generan mayor señal electroencefalográfica, es decir, mayor actividad bioeléctrica cerebral; lo cual podría implicar un aumento del arousal que conlleve, a su vez, a mayores efectos transmodales (Näätänen, 1990; Näätänen, y cols., 1978; Ruusuvirta y Astikainen, 2012).

Los estudios llevados a cabo en la década de los 70 por Deutsch y colaboradores (Deutsch, 1976, 1978; Deutsch y Roll, 1976) ya mostraron que existe cierto nivel de sesgo perceptivo en el procesamiento de frecuencias altas con respecto a frecuencias más bajas cuando éstas son presentadas conjuntamente. Es decir, el sistema auditivo ante la competencia de dos sonidos (por ejemplo, 400Hz y 800Hz) tiende a percibir preferentemente la fuente del estímulo donde se ha emitido el sonido agudo.

En nuestro Estudio 1, no se observaron efectos transmodales entre el sonido y la detección de estímulos visuales en el eje horizontal. Estos resultados van en la línea de estudios previos que sugieren que este tipo de asociación suele mostrarse sólo en músicos expertos (Rusconi y cols., 2006, Lidji y cols., 2007). Los estudios realizados con participantes sin formación musical no observaron efectos transmodales en el eje horizontal (Chiou y Rich, 2012), indicando que la asociación entre la altura tonal y la posición espacial en el eje horizontal parece ser debida a un extenso entrenamiento musical. Por lo tanto, otra posible conclusión indirecta de este resultado sería que el posible re-mapeo entre la altura tonal y la localización horizontal es menos robusto que en el eje vertical. Sin embargo, los resultados obtenidos por Lidji y colaboradores

(2007) mostraron que los efectos transmodales surgidos en el eje horizontal no son debidos a la experiencia motora, puesto que los pianistas no mostraron mayores efectos que otros instrumentistas musicales. Una posible explicación, al margen de la experiencia motora, estaría basada en la teoría de magnitud (ATOM, Walsh, 2003). La teoría de Walsh afirma que, puesto que las cantidades contables e incontables formarían parte de un sistema de magnitudes procesado por mecanismos cerebrales comunes, los fenómenos como SNARC (Dehaene, Bossini, y Giraux, 1993) y SMARC (Rusconi y cols., 2006) podrían deberse a que los números y la altura tonal compartirían un sistema común de representación mental con la ubicación espacial horizontal (por ejemplo, situar los número mayores y los tonos más agudos a la derecha del eje horizontal). Por tanto, teniendo en cuenta que los músicos tienen más experiencia categorizando magnitudes específicas como la altura tonal, ello derivaría en una estructura perceptiva que les permitiese categorizar con mayor facilidad otros grados de magnitud, como por ejemplo, la ubicación espacial en el eje horizontal (Cho, Bae, y Proctor, 2012; Walsh, 2003).

Por último, los efectos de señalización espacial que encontramos en nuestro estudio parecen no ser específicos para una posición en particular (por ejemplo, un tono ascendente y el estímulo visual arriba a la derecha), sino que este efecto parece producirse de manera general e inespecífica, abarcando un área relativamente amplia del campo visual (por encima del punto de fijación). Los resultados del estudio de Mossbridge, Grabowecky, y Suzuki (2011) revelaron que los sonidos ascendentes y descendentes solo modulaban la atención espacial si el target visual aparecía encima o debajo del punto de fijación. En cambio, si éste se situaba en una posición diagonal respecto al punto de fijación (por ejemplo, arriba-derecha) este efecto desaparecía. Tanto el Estudio 1 del presente trabajo, como el estudio llevado a cabo por Mossbridge y colaboradores (2011), sugerirían que la representación espacial de la altura tonal se produce de una manera “global” y no “local” o “específica”.

6.2 Asociaciones transmodales en bebés prelingüísticos

Una cuestión muy importante en el estudio de asociaciones transmodales es si éstas son derivadas de la experiencia con el medio o si, por lo contrario, son innatas. Varios estudios realizados con bebés prelingüísticos muestran resultados contradictorios. Por una parte, los estudios realizados por Walker y colaboradores (2010), y, más tarde, por Dolscheid y colaboradores (2014), sugieren una presencia muy temprana de asociaciones audiovisuales transmodales (en bebés de 3 meses de edad). Sin embargo, Lewkowicz y Minar (2014) no pudieron replicar los resultados del estudio de Walker y colaboradores (2010), siendo incapaces de encontrar evidencia de correspondencias transmodales entre la altura tonal y la elevación espacial (así como tampoco entre el sonido y la forma visual). Walker y colaboradores (2014) justificaron el fallo en la réplica debido a que en su estudio se utilizó variaciones en la intensidad del sonido con el objetivo de que este fuera acorde con el movimiento de la bola que se desplazaba gradualmente arriba y abajo, es decir, cuando el objeto visual se acercaba a los extremos de la pantalla la frecuencia del sonido que lo acompañaba tenía una intensidad menor. En cambio, Lewkowicz y Minar (2014) aumentaban la intensidad del sonido a medida que el sonido iba aumentando en frecuencia - hasta que la frecuencia alcanzaba los 1000Hz. Esto podría haber creado una disociación entre el movimiento del objeto visual y el sonido, anulando una posible asociación transmodal (ver Figura 13). En el Estudio 2 de la presente disertación, la intensidad de los sonidos utilizados no fue modificada en lo más mínimo (es decir, los sonidos se mantenían al mismo volumen durante todo el ensayo). Además, nuestro objeto visual no representaba movimiento en el eje vertical, sino variación de tamaño, así que la modulación de la intensidad no estaba justificada.

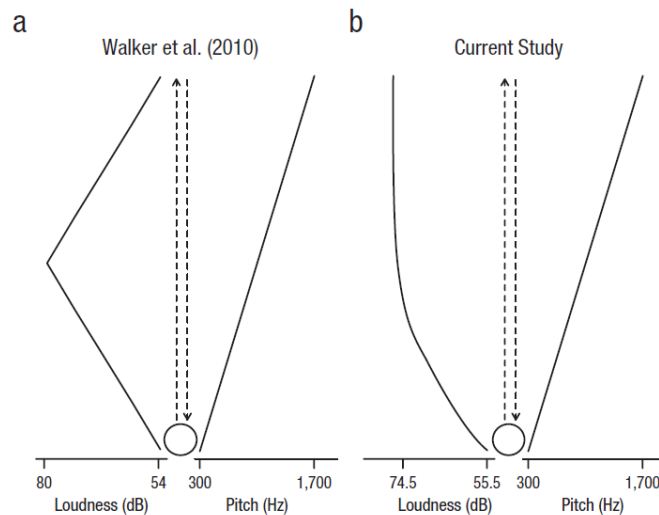


Figura 13. Variación de intensidad y la frecuencia auditiva en ensayos congruentes. (a) Patrón de variación del sonido del estudio de Walker y colaboradores (2010), a la izquierda el cambio de intensidad, a la derecha a la ascensión de la frecuencia auditiva. (b) Patrón de variación del sonido de la réplica de Lewkowicz y Minar 2014, a la izquierda el cambio de intensidad, a la derecha a la ascensión de la frecuencia auditiva (Lewkowicz y Minar, 2014).

En el Estudio 2, se investigó la posible asociación entre la frecuencia del sonido y el tamaño visual de los objetos en bebés de 4 y de 6 meses de edad. Para ello, empleamos los mismos estímulos auditivos (sonido ascendente: 300 a 1700 Hz; y sonido descendente: de 1700Hz a 300Hz) que los tres anteriores estudios mencionados. Sin embargo, nuestro estudio versaba sobre la posible relación asociativa entre altura tonal y el tamaño visual, debido a que probablemente esta asociación presenta una regularidad y robustez mayor en el medio natural que la asociación entre la altura tonal y la elevación espacial. La asociación entre altura tonal y tamaño, puede derivar directamente de la constatación física de que en el medio las frecuencias más bajas tienen mayor longitud de onda y requieren de objetos con una gran superficie vibratoria, es decir, mayor tamaño, para producir ondas más largas. Sin embargo, la asociación entre altura tonal y altura visuoespacial, presentaría un menor nivel de regularidad estadística en el medio, si bien es cierto, que los sonidos que se propagan cerca del suelo tienden a que su espectro de alta frecuencia sea absorbido por diversos objetos (generalmente existen mayor número de objetos a nivel del suelo que en zonas superiores), ello implica que si el observador no está situado lo cerca del suelo esta ley física no es aplicable.

Los resultados obtenidos mostraron que sólo los bebés de 6 meses de edad presentaban preferencia por mirar a estímulos audiovisualmente congruentes (por ejemplo, sonido ascendente con un objeto visual que disminuye de tamaño) en comparación con estímulos audiovisualmente incongruentes. Estos resultados no se observaron en bebés de 4 meses, ya que éstos no mostraron diferencias de tiempo de mirada entre las dos condiciones. Nuestros datos sugieren, por tanto, que es necesaria una mínima exposición a las regularidades del medio y/o cierto grado de maduración de los mecanismos cerebrales que sustentan la consolidación de determinadas asociaciones transmodales entre la altura tonal y el tamaño visual.

Peña, Mehler, y Nespor (2011) mostraron que los bebés prelingüísticos son capaces de asociar sonidos del habla (sílabas en este caso) con distintos tamaños de objetos. Observaron que los bebés miraban durante más tiempo a los estímulos congruentes, es decir, a un objeto visual pequeño que aparecía conjuntamente con el sonido de una sílaba que contenía una frecuencia más alta (sílaba formada con las vocales /i/ o /e/) o a un objeto visual grande que aparecía conjuntamente con una sílaba que contenía una frecuencia más baja (sílaba formada con las vocales /a/ o /o/). Los resultados del Estudio 2 demuestran que la asociación entre sonido y tamaño también está presente en sonidos que no pertenecen al habla. En este caso, no existiría una relación tan directa entre los atributos perceptivos estudiados, basada en la posible experiencia previa de los bebés como en el caso del estudio anterior: la observación o exposición del tamaño de la abertura de la boca de las personas del entorno más cercano del bebé, durante la pronunciación de distintas vocales.

La importancia del Estudio 2 radica en la edad de los participantes y su condición de bebés prelingüísticos. Si bien, en el estudio de Mondloch y Maurer (2004) se hallaron resultados similares a nuestro estudio en niños de 2,5 a 3 años de edad, es importante mencionar que estos niños mucho más mayores ya cuentan con una amplia experiencia lingüística y conocimiento de etiquetas léxicas espaciales. En relación al papel del lenguaje en las asociaciones transmodales, Melara y Marks (1990) mostraron que existen efectos transmodales, en adultos, entre la altura tonal y etiquetas léxicas espaciales. En su estudio los participantes debían discriminar las palabras "high" (alto) y "low" (bajo) que se presentaban escritas junto a un tono agudo o grave. Los datos mostraron que los participantes eran más rápidos

respondiendo a la palabra “high” cuando se presentaba con un tono agudo y la palabra “low” cuando se presentaba con un tono grave. Estos resultados sugieren que el lenguaje interfiere en las correspondencias transmodales audiovisuales.

De acuerdo con otros estudios sobre asociaciones transmodales en bebés prelingüísticos (Dolscheid y cols., 2014; Walker y cols., 2010), nuestros datos implican que el papel del lenguaje no parece ser esencial para los mismos, aunque la influencia de la adquisición de etiquetas léxicas espaciales del lenguaje podría fortalecer y/o modificar dichas asociaciones (Marks, 1999; Melara y Marks, 1990). La utilización de etiquetas léxicas, podría influir en las correspondencias transmodales, y en consecuencia, los mismos conceptos semánticos serían activados para la dimensión auditiva y visuoespacial.

Diversos estudios han sugerido que la experiencia lingüística modula varios aspectos de la cognición y la percepción (Lupyan, 2012). Una cuestión interesante sigue siendo la posibilidad de que el vínculo entre determinadas dimensiones espaciales (arriba/abajo, por ejemplo) y auditivas (como la altura tonal o la intensidad del sonido) podría estar influenciada por la activación de metáforas lingüísticas comunes, una vez que éstas ya se han adquirido (Casasanto, 2014). De hecho, la mayoría de las culturas representan simbólicamente la altura tonal verticalmente (por ejemplo, en la notación musical). Esta representación metafórica se observa también en un vocabulario común para ambas dimensiones. Por ejemplo, los términos "high" y "low" en inglés se refieren tanto a conceptos auditivos como espaciales. Sin embargo, las lenguas romances tienden a utilizar en menor medida términos espaciales para definir la altura tonal (por ejemplo, agudo y grave en español). Ello podría implicar que las lenguas que comparten términos léxicos espaciales y auditivos quizá fortalecerían, a diferencia de las lenguas que compartiesen léxico en menor medida, los vínculos asociativos entre la altura tonal y la elevación espacial (Fernández-Prieto, Spence, Pons, y Navarra, 2016).

Para una consolidación efectiva de las correspondencias transmodales debemos tener en cuenta el papel de la exposición al medio durante el desarrollo de los sistemas perceptivos. Si bien dichas correspondencias sugieren asociaciones entre diversas informaciones perceptuales aparentemente arbitrarias, en realidad existe una

constante exposición a regularidades estadísticas en el medio ambiente que pueden fortalecer asociaciones audiovisuales específicas. Por ejemplo, debido a la propia física acústica, los objetos más pequeños y más grandes tenderán a producir sonidos más agudos y más graves, respectivamente. En lo que respecta a la asociación entre altura tonal y elevación (o localización espacial), Parise, Knorre, y Ernst (2014) demostraron recientemente que los sonidos más agudos se originan generalmente a partir de fuentes posicionadas más altas, y los sonidos más graves desde posiciones más bajas en el espacio. Puesto que estas correlaciones estadísticas se derivan de la experiencia en el entorno físico real, el lenguaje no sería requisito indispensable para el desarrollo de asociaciones transmodales aunque podría modular dichas asociaciones (Lakoff y Johnson, 1980). Los bebés y los niños viven en un ambiente multisensorial basado en experiencias diarias que, combinadas con la plasticidad cerebral del sistema nervioso, puede promover la adquisición de diversas asociaciones de este tipo.

6.3 Procesamiento auditivo en pacientes con alteraciones visuoespaciales

Tanto el Estudio 1 como el Estudio 2 del presente trabajo mostraron que existen asociaciones transmodales entre determinados sonidos y determinadas características visuoespaciales, tanto en adultos como en bebés. En nuestro tercer estudio, nos planteamos si estas asociaciones podrían verse afectadas en personas con alteraciones relacionadas con el procesamiento visuoespacial. Desafortunadamente para el propósito del estudio, las lesiones aisladas en áreas parietales son poco comunes e imposibilitaron la obtención de muestra. La amplia mayoría de casos en los que pudimos observar daño parietal, no se trataba de lesiones localizadas y afectaban en mayor o menor medida a otros procesos cognitivos y perceptivos (por ejemplo, daños causados por un infarto que se extendían a la corteza auditiva).

Debido a estas limitaciones, decidimos seleccionar para este estudio pacientes con trastorno del aprendizaje no verbal (TANV). Este trastorno neurológico está

relacionado con alteraciones en el hemisferio derecho (HD) y afecta significativamente al procesamiento espacial de los estímulos externos y al rendimiento en áreas académicas como las matemáticas. Además, aunque menos descritos los pacientes con TANV también presentan problemas en habilidades sociales, así como alteraciones de la prosodia. La alteración del HD implica problemas de decodificación de la altura tonal, proceso clave en la percepción y comprensión de la prosodia (Zatorre, 1984).

En el Estudio 3, los participantes con TANV mostraron un procesamiento deficiente en la ejecución de una prueba visuoespacial en comparación con el grupo control. En dicha prueba, los participantes tenían que juzgar la dirección del movimiento de un objeto visual. Las deficiencias en el procesamiento observadas en el grupo de TANV probablemente fueron debidas a la afectación visuoespacial característica de este desorden neurológico. Además, el grupo TANV también mostró un peor desempeño al categorizar la dirección de un cambio dinámico de la altura tonal de un sonido (ascendente vs. descendente). Finalmente, se encontró, en el grupo TANV y el grupo control, una correlación positiva entre los tiempos de reacción (TRs) en la tarea de categorización visuoespacial y los TRs en la prueba de categorización de la altura tonal.

Los datos del Estudio 3 sugieren que el procesamiento visuoespacial y el procesamiento de la altura tonal podrían compartir mecanismos neuronales. Los estudios de neuroimagen de Foster y Zatorre (2010a, 2010b), en participantes sin alteraciones neurológicas, han mostrado que determinadas áreas parietales del HD especializadas en el procesamiento del espacio (por ejemplo, el surco intraparietal, IPS) muestran activación en tareas de transposición de tonos musicales. En consecuencia, las alteraciones funcionales o estructurales en el hemisferio derecho, por ejemplo, en los pacientes con TANV, podrían conllevar déficits en el procesamiento de altura tonal.

La presencia de determinados procesamientos auditivos en pacientes con TANV iría en línea con la investigación de Douglas y Bilkey (2007) con participantes amúsicos. La amusia (o sordera tonal) se caracteriza por ser un síndrome neurológico en el cual se observan importantes déficits en la percepción tonal. En el estudio de Douglas y

Bilkey (2007) los participantes amúsicos mostraron un pobre desempeño en tareas de rotación mental (Särkämö y cols., 2010; sin embargo, para resultados contradictorios véase Tillmann y cols., 2010). Las evidencias mostradas en pacientes con alteraciones neurológicas indicarían que daños específicos en áreas tradicionalmente relacionadas con el procesamiento espacial de los estímulos pueden afectar el rendimiento en tareas auditivas y viceversa. La representación de la altura tonal y la visuoespacial (por ejemplo, la rotación mental) podrían estar utilizando, en alguno de los estadios de su procesamiento, los mismos mecanismos cerebrales y cognitivos.

En el Estudio 3, los pacientes con TANV mostraron una baja velocidad de discriminación de la altura tonal, además de una elevada tasa de errores juzgando la altura tonal. Las evidencias de anomalías en la discriminación auditiva reportadas en el Estudio 3 podrían sugerir que las alteraciones en el procesamiento de la prosodia del habla en pacientes con TANV derivarían de un déficit en la discriminación de la altura tonal. Para un correcto procesamiento de la prosodia del habla, la discriminación de frecuencias auditivas es esencial (Patel y cols., 1998). Los resultados obtenidos en el Estudio 3 podrían explicar las dificultades para producir y procesar la entonación del lenguaje en estos pacientes. Los problemas de prosodia en los pacientes con TANV producen déficits en la discriminación de enunciados, por ejemplo, diferenciar entre una afirmación y una pregunta. Rourke (1988) sugirió una explicación neurológica de las alteraciones en el procesamiento de la prosodia en pacientes con TANV. Esta alteración podría deberse a anomalías estructurales de la materia blanca en el HD de estos pacientes. El HD, en concreto las regiones temporoparietales, está implicado en la percepción de la prosodia, es decir, variaciones de altura tonal en el discurso (Arimitsu y cols., 2011; Homae y cols., 2006)

Otro aspecto a tener en cuenta es que los pacientes con TANV presentan problemas de habilidades sociales. Este déficit podría ser debido a una deficiente percepción y producción de la prosodia del habla que conllevaría un déficit en comunicación social, afectando directamente las habilidades sociales de estos pacientes (Rourke, 1995).

7. Conclusions

Our perceptual system tends to associate higher-pitched sounds with upper spatial positions and smaller visual sizes, and lower pitched-sounds with lower positions and bigger visual sizes. This processes, known as crossmodal correspondences or crossmodal associations, emerges consistently and almost universally.

The cognitive, developmental and neural mechanisms underlying these crossmodal correspondences were investigated in the present doctoral dissertation. In the present work, I studied how the crossmodal associations between auditory and visuospatial stimuli modulates cognitive processes in infancy and maturity. In addition, I investigated the impairments in auditory processing in population with visuospatial disorders.

The presence of crossmodal associations between pitch and spatial elevation are solid in adults (see Spence, 2011, for a review), and can be observed in the ability of certain auditory stimuli to modulate visuospatial attention (e.g., rising sounds facilitate the detection of a visual stimuli in upper areas in the external space). However, it seems that higher-pitched sounds or sounds with rising frequencies have higher inherent spatial properties than lower-pitched or falling frequency sounds. These spatial properties of the sound could modulate the perceptual system and facilitate the detection of visual objects in superior positions of the visual field.

We found evidence of the existence of crossmodal correspondences between auditory and visuospatial dimensions also in prelinguistic infants, at the age of 6 months. We observed that 4-month-old infants have not developed yet the mechanisms that facilitate audiovisual associations. This pattern of results suggests that the experience with the physical world and/or further maturation is needed to fully develop certain audiovisual crossmodal associations.

Finally, in Study 3, I observed that patients with NLD, that show impairments in visuospatial skills, show a deficit in pitch-related tasks. These evidences could suggest that auditory and visuospatial processes share common mechanisms in some stages

of the perceptual processing, that is, for example, that the impairment in visuospatial skills could disrupt the auditory processing.

8. Futuras direcciones

A través de los diferentes estudios presentados en esta disertación hemos podido observar como el fenómeno perceptivo de las correspondencias transmodales modula nuestras respuestas a diferentes estímulos sensoriales. Estos hallazgos podrían tener un impacto en diversas disciplinas. Desde el punto de vista del estudio del procesamiento de la integración entre la altura tonal y determinados atributos visuoespaciales (por ejemplo, el tamaño visual o la ubicación espacial), en futuras investigaciones sería interesante incorporar el uso de otras técnicas, como la neuroimagen, que podría arrojar luz sobre el desarrollo de los mecanismos neurales implicados en las correspondencias transmodales y el papel de otros sistemas cognitivos, por ejemplo, la activación de determinadas áreas cerebrales asociadas con el procesamiento visuoespaciales durante tareas relacionadas con la altura tonal.

A nivel del estudio del desarrollo cognitivo de la infancia y la niñez, sería interesante evaluar, en futuros estudios, cómo las asociaciones transmodales se van consolidando debido a la madurez y/o la exposición al medio, viéndose, a su vez, moduladas por el desarrollo del lenguaje y/o a la adquisición de etiquetas léxicas auditivas y visuoespaciales.

Por último, desde el enfoque clínico, los resultados de esta disertación doctoral pueden abrir nuevos caminos en el diagnóstico y tratamiento de los trastornos que afectan al procesamiento visuoespacial. Por ejemplo, resulta primordial, según la óptica desarrollada en el presente trabajo, investigar la posible incorporación de pruebas neuropsicológicas que evalúen de una forma detallada la percepción de la altura tonal y la comprensión de la prosodia en la evaluación de trastornos visuoespaciales de base neurológica (por ejemplo, en el TANV o lesiones en el lóbulo parietal derecho).

9. General references

- Arimitsu, T., Uchida-Ota, M., Yagihashi, T., Kojima, S., Watanabe, S., Hokuto, I., . . . Minagawa-Kawai, Y. (2011). Functional hemispheric specialization in processing phonemic and prosodic auditory changes in neonates. *Frontiers in Psychology, 2*.
- Baron-Cohen, S., Burt, L., Smith-Laittan, F., Harrison, J., & Bolton, P. (1996). Synaesthesia: prevalence and familiarity. *Perception, 25*(9), 1073-1079. doi: 10.1068/p251073
- Baron-Cohen, S., Harrison, J., Goldstein, L. H., & Wyke, M. (1993). Coloured speech perception: Is synaesthesia what happens when modularity breaks down?. *Perception, 22*(4), 419-426. doi: 10.1068/p220419
- Bayne, T., & Spence, C. (2015) Multisensory perception. In M. Matthen (Ed.). *The Oxford Handbook of Philosophy of Perception* (pp. 603-620) Oxford University Press, USA
- Belkin, K., Martin, R., Kemp, S. E., & Gilbert, A. N. (1997). Auditory pitch as a perceptual analogue to odor quality. *Psychological Science, 8*(4), 340-342. doi: 10.1111/j.1467-9280.1997.tb00450.x
- Ben-Artzi, E., & Marks, L. (1995). Visual-auditory interaction in speeded classification: Role of stimulus difference. *Attention, Perception, & Psychophysics, 57*(8), 1151-1162. doi: 10.3758/bf03208371
- Bernstein, I. H., & Edelman, B. A. (1971). Effects of some variations in auditory input upon visual choice reaction time. *Journal of experimental psychology, 87*(2), 241-247. doi: 10.1037/h0030524
- Bertelson, P., & de Gelder, B. (2004). The psychology of multimodal perception. In C. Spence & J. Driver (Eds.), *Crossmodal space and crossmodal attention* (pp. 141–177). Oxford: Oxford University Press.
- Bever, T. G., & Chiarello, R. J. (1974). cerebral dominance in musicians and nonmusicians. *Science, 185*(4150), 537-539. doi: 10.1126/science.185.4150.537
- Bien, N., ten Oever, S., Goebel, R., & Sack, A. T. (2012). The sound of size: crossmodal binding in pitch-size synesthesia: a combined TMS, EEG and psychophysics study. *Neuroimage, 59*(1), 663-672. doi: 10.1016/j.neuroimage.2011.06.095

- Calvert, G., Spence, C., & Stein, B. E. (2004). *The handbook of multisensory processes*. MIT press.
- Casasanto, D. (2014). Development of metaphorical thinking: The role of language. In *Language and the creative mind*. M. Borkent, J. Hinell, & B. Dancygier (Eds.) Stanford, CA: CSLI Publications.
- Chiou, R., & Rich, A. N. (2012). Cross-modality correspondence between pitch and spatial location modulates attentional orienting. *Perception, 41*(3), 339-353.
- Cho, Y. S., Bae, G. Y., & Proctor, R. W. (2012). Referential coding contributes to the horizontal SMARC effect. *Journal of Experimental Psychology: Human Perception and Performance, 38*(3), 726-734. doi: 10.1037/a0026157
- Cohen Kadosh, R., Bien, N., & Sack, A. T. (2012). Automatic and intentional number processing both rely on intact right parietal cortex: a combined fMRI and neuronavigated TMS study. *Frontiers in human neuroscience, 6*, 2. doi: 10.3389/fnhum.2012.00002
- Cohen Kadosh, R., Cohen Kadosh, K., & Henik, A. (2007). The neuronal correlate of bidirectional synesthesia: A combined event-related potential and functional magnetic resonance imaging study. *Journal of Cognitive Neuroscience, 19*(12), 2050-2059. doi: 10.1162/jocn.2007.19.12.2050
- Crisinel, A. S., & Spence, C. (2009). Implicit association between basic tastes and pitch. *Neuroscience letters, 464*(1), 39-42. doi: 10.1016/j.neulet.2009.08.016
- Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General, 122*(3), 371-396. doi: 10.1037/0096-3445.122.3.371371-396.
- Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General, 122*(3), 371-396. doi: 10.1037/0096-3445.122.3.371
- Deroy, O., & Spence, C. (2013). Why we are not all synesthetes (not even weakly so). *Psychonomic bulletin & review, 20*(4), 643-664. doi: 10.3758/s13423-013-0387-2
- Deroy, O., Fernandez-Prieto, I., Navarra, J., & Spence, C. (in press). Unravelling the paradox of spatial pitch. To appear in T. L. Hubbard (Eds.), *Spatial biases in perception and cognition*. Cambridge, UK: Cambridge University Press

- Deutsch, D. (1976). Lateralization by frequency in dichotic tonal sequences as a function of interaural amplitude and time differences. *The Journal of the Acoustical Society of America*, 60(S1), S50-S50. doi: 10.1121/1.2003382
- Deutsch, D. (1978). Lateralization by frequency for repeating sequences of dichotic 400- and 800-Hz tones. *The Journal of the Acoustical Society of America*, 63(1), 184-186. doi: 10.1121/1.381710
- Deutsch, D., & Roll, P. L. (1976). Separate "what" and "where" decision mechanisms in processing a dichotic tonal sequence. *Journal of Experimental Psychology: Human Perception and Performance*, 2(1), 23-29. doi: 10.1037/0096-1523.2.1.23
- Dolscheid, S., Hunnius, S., Casasanto, D., & Majid, A. (2014). Prelinguistic infants are sensitive to space-pitch associations found across cultures. *Psychological Science*, 25(6), 1256-1261. doi: 10.1177/0956797614528521
- Douglas, K. M., & Bilkey, D. K. (2007). Amusia is associated with deficits in spatial processing. *Nature neuroscience*, 10(7), 915-921. doi: http://www.nature.com/neuro/journal/v10/n7/supinfo/n1925_S1.html
- Ernst, M. O., & Bühlhoff, H. H. (2004). Merging the senses into a robust percept. *Trends in cognitive sciences*, 8(4), 162-169. doi: 10.1016/j.tics.2004.02.002
- Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of vision*, 10(1), 6-6. doi: 10.1167/10.1.6
- Fletcher, H., & Munson, W. A. (1933). Loudness, its definition, measurement and calculation. *Bell System Technical Journal*, 12(4), 377-430. doi: 10.1121/1.1915637
- Fort, M., Weiß, A., Martin, A., & Peperkamp, S. (2013, August). Looking for the bouba-kiki effect in prelexical infants. In *AVSP* (pp. 71-76).
- Foster, N. E., & Zatorre, R. J. (2010a). Cortical structure predicts success in performing musical transformation judgments. *Neuroimage*, 53(1), 26-36. doi: 10.1016/j.neuroimage.2010.06.042
- Foster, N. E., & Zatorre, R. J. (2010b). A role for the intraparietal sulcus in transforming musical pitch information. *Cerebral Cortex*, 20(6), 1350-1359. doi: 10.1093/cercor/bhp199

- Gallace, A., & Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of visual size. *Attention, Perception, & Psychophysics*, *68*(7), 1191-1203. doi: 10.3758/bf03193720
- Grefkes, C., & Fink, G. R. (2005). The functional organization of the intraparietal sulcus in humans and monkeys. *Journal of Anatomy*, *207*(1), 3-17. doi: 10.1111/j.1469-7580.2005.00426.x
- Homae, F., Watanabe, H., Nakano, T., Asakawa, K., & Taga, G. (2006). The right hemisphere of sleeping infant perceives sentential prosody. *Neuroscience Research*, *54*(4), 276-280. doi: 10.1016/j.neures.2005.12.006
- Humphries, C., Liebenthal, E., & Binder, J. R. (2010). Tonotopic organization of human auditory cortex. *Neuroimage*, *50*(3), 1202-1211. doi: 10.1016/j.neuroimage.2010.01.046
- Husain, M., & Nachev, P. (2007). Space and the parietal cortex. *Trends in cognitive sciences*, *11*(1), 30-36. doi: <http://dx.doi.org/10.1016/j.tics.2006.10.011>
- Johnsrude, I. S., Penhune, V. B., & Zatorre, R. J. (2000). Functional specificity in the right human auditory cortex for perceiving pitch direction. *Brain*, *123* (Pt 1), 155-163.
- Kaasinen, S., Pyykko, I., Ishizaki, H., & Aalto, H. (1995). Effect of intratympanically administered gentamicin on hearing and tinnitus in Meniere's disease. *Acta Oto-Laryngologica*, *520 Pt 1*, 184-185.
- King, A. J., & Palmer, A. R. (1985). Integration of visual and auditory information in bimodal neurones in the guinea-pig superior colliculus. *Experimental Brain Research*, *60*(3), 492-500. doi: 10.1007/bf00236934
- Kishon-Rabin, L., Roth, D. A., Van Dijk, B., Yinon, T., & Amir, O. (2004). Frequency discrimination thresholds: the effect of increment versus decrement detection of frequency. *Journal of basic and clinical physiology and pharmacology*, *15*(1-2), 29-40. doi: 10.1515/JBCPP.2004.15.1-2.29
- Köhler, W. (1929), *Gestalt Psychology*. (New York: Liveright).
- Kong, J., Wang, C., Kwong, K., Vangel, M., Chua, E., & Gollub, R. (2005). The neural substrate of arithmetic operations and procedure complexity. *Cognitive Brain Research*, *22*(3), 397-405. doi: 10.1016/j.cogbrainres.2004.09.011

- Lakoff, G., & Johnson, M. (1980). The metaphorical structure of the human conceptual system. *Cognitive science*, 4(2), 195-208. doi: 10.1016/S0364-0213(80)80017-6
- Lewkowicz, D. J., & Minar, N. J. (2014). Infants are not sensitive to synesthetic cross-modality correspondences: A comment on Walker et al. (2010). *Psychological Science*. doi: 10.1177/0956797613516011
- Lidji, P., Kolinsky, R., Lochy, A., & Morais, J. (2007). Spatial associations for musical stimuli: A piano in the head? *Journal of Experimental Psychology-Human Perception and Performance*, 33(5), 1189-1207.
- Liégeois-Chauvel, C., Peretz, I., Babai, M., Laguitton, V., & Chauvel, P. (1998). Contribution of different cortical areas in the temporal lobes to music processing. *Brain*, 121(10), 1853-1867. doi: 10.1093/brain/121.10.1853
- Lupyan, G. (2012). Linguistically modulated perception and cognition: The label-feedback hypothesis. *Frontiers in Psychology*, 3(54). doi: 10.3389/fpsyg.2012.00054
- Marks, L. E., Hammeal, R. J., & Bornstein, M. H. (1987). Perceiving similarity and comprehending metaphor. *Monographs of the Society for Research in Child Development*, 52(1), 1-102.
- Martino, G., & Marks, L. E. (2000). Cross-modal interaction between vision and touch: the role of synesthetic correspondence. *Perception*, 29(6), 745-754. doi: 10.1068/p2984
- Martino, G., & Marks, L. E. (2001). Synesthesia: strong and weak. *Current Directions in Psychological Science*, 10(2), 61-65. doi: 10.1111/1467-8721.00116
- Maurer, D., Gibson, L. C., & Spector, F. (2012). Infant synaesthesia. *Multisensory development*, 229.
- Maurer, D., Pathman, T., & Mondloch, C. J. (2006). The shape of boubas: Sound-shape correspondences in toddlers and adults. *Developmental science*, 9(3), 316-322. doi: 10.1111/j.1467-7687.2006.00495.x
- Melara, R. D., & Marks, L. E. (1990). Processes underlying dimensional interactions: Correspondences between linguistic and nonlinguistic dimensions. *Memory & Cognition*, 18(5), 477-495. doi: 10.3758/BF03198481

- Melara, R. D., & O'Brien, T. P. (1987). Interaction between synesthetically corresponding dimensions. *Journal of Experimental Psychology: General*, *116*(4), 323-336. doi: 10.1037/0096-3445.116.4.323
- Miller, J. (1991). Channel interaction and the redundant-targets effect in bimodal divided attention. *Journal of Experimental Psychology: Human Perception and Performance*, *17*(1), 160-169. doi: 10.1037/0096-1523.17.1.160
- Milner, B. (1962). Laterality effects in audition. *Interhemispheric relations and cerebral dominance*, 177-195.
- Mondloch, C., & Maurer, D. (2004). Do small white balls squeak? Pitch-object correspondences in young children. *Cognitive, Affective, & Behavioral Neuroscience*, *4*(2), 133-136. doi: 10.3758/CABN.4.2.133
- Mossbridge, J. A., Grabowecky, M., & Suzuki, S. (2011). Changes in auditory frequency guide visual-spatial attention. *Cognition*, *121*(1), 133-139. doi: 10.1016/j.cognition.2011.06.003
- Mudd, S. A. (1963b). Spatial stereotypes of four dimensions of pure tone. *Journal of experimental psychology*, *66*(4), 347-352. doi: 10.1037/h0040045
- Näätänen, R. (1990). The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behavioral and Brain Sciences*, *13*(02), 201-233. doi: doi:10.1017/S0140525X00078407
- Näätänen, R., Gaillard, A. W. K., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica (Amst)*, *42*(4), 313-329. doi: 10.1016/0001-6918(78)90006-9
- Naumer, M. J., & van den Bosch, J. J. (2009). Touching sounds: thalamocortical plasticity and the neural basis of multisensory integration. *Journal of neurophysiology*, *102*(1), 7-8. doi: 10.1152/jn.00209.2009
- Occelli, V., Spence, C., & Zampini, M. (2009). Compatibility effects between sound frequency and tactile elevation. *Neuroreport*, *20*(8), 793-797. doi: 10.1097/WNR.1090b1013e32832b38069.
- Olsho, L. W. (1984). Infant frequency discrimination. *Infant behavior and development*, *7*(1), 27-35. doi: 10.1016/S0163-6383(84)80020-X

- Olsho, L. W., Koch, E. G., & Halpin, C. F. (1987). Level and age effects in infant frequency discrimination. *The Journal of the Acoustical Society of America*, *82*(2), 454-464. doi: 10.1121/1.395446
- Ozturk, O., Krehm, M., & Vouloumanos, A. (2013). Sound symbolism in infancy: evidence for sound–shape cross-modal correspondences in 4-month-olds. *Journal of experimental child psychology*, *114*(2), 173-186. doi: 10.1016/j.jecp.2012.05.004
- Parise, C. V., Knorre, K., & Ernst, M. O. (2014). Natural auditory scene statistics shapes human spatial hearing. *Proceedings of the National Academy of Sciences*, *111*(16), 6104-6108.
- Parkinson, C., Kohler, P. J., Sievers, B., & Wheatley, T. (2012). Associations between auditory pitch and visual elevation do not depend on language: evidence from a remote population. *Perception-London*, *41*(7), 854. doi:10.1068/p7225
- Patching, G. R., & Quinlan, P. T. (2002). Garner and congruence effects in the speeded classification of bimodal signals. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(4), 755-775. doi: 10.1037/0096-1523.28.4.755
- Pearce, M. T. (2005). The construction and evaluation of statistical models of melodic structure in music perception and composition. London: City University.
- Peretz, I. (1990). Processing of local and global musical information by unilateral brain-damaged patients. *Brain*, *113*(4), 1185-1205. doi: 10.1093/brain/113.4.1185
- Peretz, I., & Hyde, K. L. (2003). What is specific to music processing? Insights from congenital amusia. *Trends in cognitive sciences*, *7*(8), 362-367. doi: 10.1016/S1364-6613(03)00150-5
- Peretz, I., & Zatorre, R. J. (2005). Brain organization for music processing. *Annu Rev Psychol*, *56*, 89-114. doi: 10.1146/annurev.psych.56.091103.070225
- Peretz, I., & Zatorre, R. J. (Eds.). (2003). *The cognitive neuroscience of music* (pp. 192-203). Oxford: Oxford University Press.
- Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*(1), 3-25.

- Pratt, C. C. (1930). The spatial character of high and low tones. *J. Exp. Psychol.*, *13*, 278-285.
- Proctor, R. W., & Reeve, T. G. (Eds.). (1989). *Stimulus-response compatibility: An integrated perspective*. Elsevier.
- Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia -- A window into perception, thought and language. *Journal of Consciousness Studies*, *8*(12), 3-34.
- Rauscher, F. H., Shaw, G. L., & Ky, K. N. (1993). Music and spatial task performance. *Nature*, *365*, 611.
- Rizzo, M., & Eslinger, P. J. (1989). Colored hearing synesthesia An investigation of neural factors. *Neurology*, *39*(6), 781-781. doi: 10.1212/WNL.39.6.781
- Roffler, S. K., & Butler, R. A. (1968). Localization of Tonal Stimuli in the Vertical Plane. *J Acoust Soc Am*, *43*(6), 1260-1266. doi: 10.1121/1.1910977
- Rusconi, E., Kwan, B., Giordano, B. L., Umiltà, C., & Butterworth, B. (2006). Spatial representation of pitch height: the SMARC effect. *Cognition*, *99*, 113-129. doi: 10.1016/j.cognition.2005.01.004
- Ruusuvirta, T. T., & Astikainen, P. (2012). Mismatch negativity of higher amplitude for melodic ascendance than descendance. *Neuroreport*, *23*(4), 220-223.
- Salgado-Montejo, A., Marmolejo-Ramos, F., Alvarado, J. A., Arboleda, J. C., Suarez, D. R., & Spence, C. (2016). Drawing sounds: representing tones and chords spatially. *Experimental Brain Research*, 1-14. doi:10.1007/s00221-016-4747-9
- Sapir, E. (1929). The status of linguistics as a science. *Language*, *5*(4), 207-214. doi: 10.2307/409588
- Särkämö, T., Tervaniemi, M., Soinila, S., Autti, T., Silvennoinen, H. M., Laine, M., ... & Pihko, E. (2010). Auditory and cognitive deficits associated with acquired amusia after stroke: a magnetoencephalography and neuropsychological follow-up study. *PLoS One*, *5*(12), e15157. doi: 10.1371/journal.pone.0015157
- Schellenberg, E. (2005). Music and cognitive abilities. *Curr. Dir. Psychological Science.*, *14*, 317-320.
- Sidtis, J. J. (1985). Bilateral language and commissurotomy. In A. G. Reeves (Ed.), *Epilepsy and the Corpus Callosum* (pp. 369-380). Boston, MA: Springer US.

- Sloboda, J. A., Wise, K. J., & Peretz, I. (2005). Quantifying tone deafness in the general population. *Annals of the New York Academy of Sciences*, 1060(1), 255-261. doi: 10.1196/annals.1360.018
- Sonnadara, R. R., Gonzalez, D. A., Hansen, S., Elliott, D., & Lyons, J. L. (2009). Spatial properties of perceived pitch. *Annals of the New York Academy of Sciences*, 1169(1), 503-507. doi: 10.1111/j.1749-6632.2009.04858.x
- Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, 73(4), 971-995. doi: 10.3758/s13414-010-0073-7
- Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. The MIT Press.
- Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nature Reviews Neuroscience*, 9(4), 255-266.
- Stumpf, C. (1883). *Tonpsychologie I* [Psychology of the tone]. Leipzig: Hirzel.
- Tillmann, B., Jolicœur, P., Ishihara, M., Gosselin, N., Bertrand, O., Rossetti, Y., & Peretz, I. (2010). The amusic brain: Lost in music, but not in space. *PLoS One*, 5(4), e10173. doi: 10.1371/journal.pone.0010173
- Tomatis, A.A. (1978) *Presented in International Kodaly Symposium in Paris*.
- Tramo, M. J., Shah, G. D., & Braida, L. D. (2002). Functional role of auditory cortex in frequency processing and pitch perception. *Journal of Neurophysiology*, 87(1), 122-139.
- Trehub, S. E., Schneider, B. A., & Endman, M. (1980). Developmental changes in infants' sensitivity to octave-band noises. *Journal of Experimental Child Psychology*, 29(2), 282-293.
- Trommershäuser, J., Kording, K., & Landy, M. S. (Eds.). (2011). *Sensory cue integration*. Oxford University Press.
- von Békésy, G. (1963). Three experiments concerned with pitch perception. *Journal of the Acoustical Society of America*, 35(4), 602-606. doi: 10.1121/1.1918545
- Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A., & Johnson, S. P. (2010). Preverbal infants' sensitivity to synaesthetic cross-modality correspondences. *Psychological Science*, 21(1), 21-25. doi: 10.1177/0956797609354734
- Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A., & Johnson, S. P. (2014). Preverbal infants are sensitive to cross-sensory

- correspondences: much ado about the null results of Lewkowicz and Minar (2014). *Psychological Science*. doi: 10.1177/0956797613520170
- Walsh, V. (2003). A theory of magnitude: common cortical metrics of time, space and quantity. *Trends in cognitive sciences*, 7(11), 483-488. doi: 10.1016/j.tics.2003.09.002
- Zatorre, R. J. (1984). Musical perception and cerebral function: A critical review. *Music Perception: An Interdisciplinary Journal*, 2(2), 196-221. doi: 10.2307/40285291
- Zatorre, R. J. (1985). Discrimination and recognition of tonal melodies after unilateral cerebral excisions. *Neuropsychologia*, 23(1), 31-41. doi: 10.1016/0028-3932(85)90041-7
- Zatorre, R. J. (1988). Pitch perception of complex tones and human temporal-lobe function. *The Journal of the Acoustical Society of America*, 84(2), 566-572.
- Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: music and speech. *Trends in cognitive sciences*, 6(1), 37-46. doi: 10.1016/S1364-6613(00)01816-7
- Zatorre, R. J., Evans, A. C., & Meyer, E. (1994). Neural mechanisms underlying melodic perception and memory for pitch. *The Journal of Neuroscience*, 14(4), 1908-1919.
- Zatorre, R., Evans, A., & Meyer, E. (1994). Neural mechanisms underlying melodic perception and memory for pitch. *The Journal of Neuroscience*, 14(4), 1908-1919.

